Bildverarbeitung

Volker Aurich

Heinrich-Heine-Universität Düsseldorf

Sommersemester 2011

Inhaltsverzeichnis

0 Einleitung			1	
1	Grundlagen der digitalen Bildverarbeitung			
	1.1	Bildsignale	3	
	1.2	Diskretisierung eines Kamerabildes	4	
	1.3	Quantisierung	5	
	1.4	Das Histogramm eines Bildes	5	
	1.5	Aufbau eines Bildverarbeitungssystems für Grauwertbilder	5	
	1.6	Strategien der Bilderkennung	6	
2	Tra	nslationsinvariante, lokale Operatoren im Ortsbereich	9	
	2.1	Vereinbarung über den Definitionsbereich	9	
	2.2	Vereinbarung über den Wertebereich	10	
	2.3	Punktoperatoren	10	
	2.4	Lineare Operatoren	11	
	2.5	Nichtlineare Gradientenoperatoren	14	
	2.6	Baryzentrische Kantenoperatoren	15	
	2.7	Rangordnungsoperatoren	15	
	2.8	Morphologische Operatoren	16	
3	Bilo	lverbesserung	19	
	3.1	Verminderung von Rauschen	19	
	3.2	Hervorhebung von Helligkeitskanten	23	
	3.3	Kantenerhaltende Glättung	26	
	3.4	Hervorhebung von Linien durch Verdünnung	31	
4	Det	ektion von Kanten	33	
	4.1	Kontinuierliche Modellierung von Kanten	33	
	4.2	Diskretisierung	34	
	4.3	Cannys Kantenfinder	34	
	4.4	Marr-Hildreth-Kantenfinder	35	
	4.5	Bemerkungen	35	
	4.6	Nichtlineare Differenzenfilter mit Nulldurchgangsdetektoren	35	

INHALTSVERZEICHNIS

	4.7	Vorschalten einer kantenerhaltenden Glättung	36			
5	Bin	Binärisierung und Merkmalsextraktion 3'				
	5.1	Harte und weiche Entscheidungen	37			
	5.2	Grauwertsegmentierung und Glättung	38			
	5.3	Lokale Binärisierung kleiner Strukturen	39			
	5.4	Detektion von Linien	39			
6	Ver	Verarbeitung von Binärbildern				
	6.1	Topologische Eigenschaften von Binärbildern	45			
	6.2	Geometrische Eigenschaften zusammenhängender Binärobjekte	47			
	6.3	Bestimmung der Lage und Position von Binärobjekten	47			
	6.4	Konturverfolgung	50			
	6.5	Skelettierung	51			
	6.6	Pixelweise logische und morphologische Operationen	54			
	6.7	Double thresholding und feature-AND	54			
	6.8	Trennungsalgorithmen, der Wasserscheide-Algorithmus	54			
	6.9	Beschreibung der Gestalt	56			
7	Beschreibung linearer Filter im Frequenzbereich					
	7.1	Die Fouriertransformation	61			
	7.2	Faltungsoperatoren	66			
	7.3	Beschreibung von Faltungsoperatoren im Frequenzbereich	69			
	7.4	Bemerkungen zur Konstruktion diskreter Filter	74			
8	Anv	wendungen	77			
	8.1	Inverse Filterung	77			
	8.2	Korrelationstechniken	79			
	8.3	Die Radontransformation	80			
9	Seg	mentierung und Klassifikation	83			
	9.1	Beispielkriterien	83			
	9.2	Homogenitätskriterien	83			
	9.3	Beispiele für lokale Merkmalsvektoren	83			
	9.4	Segmentierungsstrategien	84			
	9.5	Clusteranalyse	85			
	9.6	Klassifikation	85			
	9.7	Bemerkungen	85			
\mathbf{A}	Anv	wendungsbeispiele	87			
	A.1	Richtungsabhängige morphologische Erosion	87			
	A.2	Double thresholding und feature-AND	88			
	A.3	Wasserscheide und EDA	89			

INHALTSVERZEICHNIS

В	Mat	Iathematische Ergänzung 9			
	B.1	Metrische Räume, stetige Abbildungen	91		
	B.2	Normierte Räume	92		
	B.3	Euklidische Vektorräume, Skalarprodukt	92		
	B.4	Unitäre Vektorräume, Hermitesches Skalarprodukt	93		
	B.5	Orthonormalbasen	94		
	B.6	Maßräume, Wahrscheinlichkeitsmaße \hdots	95		
	B.7	Messbare Abbildungen, Zufallsvariablen, Verteilungen, Dichten $\ \ldots \ $	96		
	B.8	Integrale	98		
	B.9	Produkträume, stochastische Unabhängigkeit $\hdotspace{1.5}$	100		
Lit	Literaturverzeichnis 103				

Literaturverzeichnis

v

INHALTSVERZEICHNIS

Kapitel 0

Einleitung

 $\label{eq:Bild} \mbox{Bild} = \mbox{zweidimensionales Signal} = \mbox{Funktion zweier Variablen mit Werten in einem m-dimensionalen Raum W}.$ **Beispiele.**

- Kamerabilder
 - Grauwertbild: Helligkeitsverteilung auf der Bildebene, in der sich der Sensor befindet, also m = 1, W Intervall in \mathbb{R} .
 - Farbbild: meist die Intensitätsverteilung des auf die Bildebene fallenden Lichts in drei verschiedenen Frequenzbereichen (Rot, Grün, Blau), also m = 3.
- Bildgebende Verfahren in der Medizin: Die Bilder werden meist erst algorithmisch mit einem Rechner aus den Daten anderer Sensoren erzeugt: Sonografie (Ultraschall), Computertomografie (Röntgenstrahlung), Kernspintomografie (Radiowellen), Emissionstomografie (radioaktive Strahlung), Elektronenmikroskopie (Elektronenstrahlung). Im allgemeinen werden Grauwertbilder erzeugt.
- Radioastronomie: Radioempfänger, Radar.

Die Bildverarbeitung beschäftigt sich mit Verfahren zur Veränderung von Bildsignalen meist mit folgenden Zielen:

- Bildverbesserung (Image enhancement), d.h. deutlichere Hervorhebung wichtiger Strukturen.
- Bildrekonstruktion, d.h. Eliminierung von Störungen, wie Rauschen, geometrische Verzerrungen und Verschmierungen durch Bewegung.
- Merkmalsextraktion, d.h. Segmentierung und eventuell Klassifizierung gewisser Bildstrukturen, wie Kanten, Muster und Objekte.
- Kompression von Bilddaten, d.h. möglichst kompakte Codierung von Bildern ohne Informationsverlust oder auch mit Verlust von Information, die aber in der beabsichtigten Anwendung nicht relevant ist.

Weiter gesteckte Ziele verfolgt Maschinelles Sehen (*Computer vision*), nämlich die Analyse von 3-dimensionalen Szenen und Bewegungen. Die Anwendungsbereiche der Bildverarbeitung sind vielfältig.

- Medizin: Tomographie u.ä., Histologie
- Strukturbiologie: Rekonstruktion der Gestalt von Makromolekülen aus Elektronenmikroskopbildern
- Robotik: berührungslose Sensorik
- Produktionstechnik: zerstörungsfreie Werkstoffprüfung, Sortieren, Kontrolle
- Sicherheitstechnik: Auswertung von Fingerabdrücken, Erkennung von Gesichtern, Augenretina

- Fernerkundung: Auswertung von multispektralen Satellitenbildern oder von Luftbildern
- Videotechnik: Datenkompression, z.B. JPEG oder MPEG.
- Militärische Anwendungen: alles

Kapitel 1

Grundlagen der digitalen Bildverarbeitung

1.1 Bildsignale

Unter einem diskreten Bild versteht man eine Abbildung $f : R \to W$ mit $R \subset \mathbb{Z}^2$ und $W \subset \mathbb{R}^m$. Gleiches gilt für ein kontinuierliches Bild, wobei $R \subset \mathbb{R}^2$ ist. Meist ist R ein Rechteck und m = 1; dann kann f als Helligkeitsverteilung veranschaulicht werden. Man spricht von einem Grauwertbild.

Im Rechner muß man den Wertebereich quantisieren, das heißt, durch einen endlichen Wertebereich W ersetzen. Für Grauwertbilder wählt man häufig $W = [\![0, N-1]\!] = \{0, \ldots, N-1\}$ mit $N = 2^n$. Vereinbarung (über die Darstellung von Grauwertbildern).

Der Wert 0 entspricht schwarz, der Wert N-1 weiß; die dazwischen liegenden Werte entsprechen monoton ansteigenden Helligkeiten, sogenannten *Grauwertstufen*.



Abbildung 1.1: Graukeil mit linearem Verlauf zwischen 0 und N-1

Vereinbarung. Wenn nichts anderes gesagt wird, ist im folgenden unter einem Grauwertbild der Größe $N_x \times N_y$ stets eine Abbildung

$$f : [0, N_x - 1] \times [0, N_y - 1] \to [0, 255]$$

mit $N_x, N_y \in \mathbb{N}$ zu verstehen.

Vereinbarung (über die Wahl der Koordinaten in einem diskreten Bild). Die Koordinaten in einem (rechteckigen) Bild seien, wie in der folgenden Abbildung dargestellt, gewählt.



Abbildung 1.2: Darstellung auf dem Monitor

Ein Bild $f: R \to W$ heißt *Binärbild*, wenn $\#f(R) \leq 2$. Die beiden Werte wird man i.a. durch schwarz und weiß darstellen. Die Punkte des Definitionsbereichs R eines Bildes heißen *Pixel (picture element)* oder *Bildpunkte*.

Bezeichnung. Für $R \subset \mathbb{Z}^2$ und $W \subset \mathbb{R}$ sei $\mathbb{B}(R, W) := \{f : R \to W\}$ die Menge der diskreten Bilder mit Definitionsbereich R und Wertebereich W. \mathbb{B} bezeichne die Menge aller partiell definierten Abbildungen $f : \mathbb{Z}^2 \xrightarrow{p} W$; man nennt sie die Menge aller diskreten, reellen Bilder.

Bemerkungen. In manchen Anwendungen ist der Wertebereich W größer, z.B. $W = [0, 2^{16} - 1]$ in vielen medizinischen Anwendungen; allerdings werden oft nur 12 Bit ausgenutzt, das heißt, die Werte von f liegen in dem Intervall $[0, 2^{12} - 1]$.

1.2 Diskretisierung eines Kamerabildes

In üblichen Videokameras wird durch ein Linsensystem auf der Bildebene eine Intensitätsverteilung (farbig) erzeugt, also ein kontinuierliches Bild. In der Bildebene sind (bei den üblichen CCD-Kameras) lichtempfindliche Zellen in einem zweidimensionalem Feld (oder auch als lineare Zeile z.B. in Scannern) angeordnet. In jeder dieser Zellen sammelt sich während der Belichtung eine Ladungsmenge an, die in etwa der Lichtintensität proportional ist. Diese Ladungen werden zeilenweise von links nach rechts bzw. von oben nach unten nacheinander in Spannungswerte umgewandelt und als eindimensionale Folge ausgegeben.

Die Diskretisierung des kontinuierlichen Bildsignals findet also in der Kamera statt! Daß das entstehende diskrete Bildsignal serialisiert wird, hat keine prinzipielle, sondern nur eine technische Bedeutung.

Wie gut wird das kontinuierliche Bildsignal durch das diskrete repräsentiert? Eine theoretische Antwort gibt Shannons Abtasttheorem. Sei $f : \mathbb{R}^2 \to \mathbb{R}$ Lebesgue-integrierbar und stetig und sei

$$\hat{f}(\xi,\eta) = \iint f(x,y) \, e^{-2\pi i x \xi} \, e^{-2\pi i y \eta} \, dx dy$$

die Fouriertransformierte von f. Verschwindet \hat{f} außerhalb des Rechtecks $\{(\xi, \eta) \mid |\xi| < \frac{1}{2\Delta x}, |\eta| < \frac{1}{2\Delta y}\}$ (Nyquist-Bedingung), so kann f aus den Abtastwerten $f(k\Delta x, l\Delta y)$ für $k, l \in \mathbb{Z}$ (mit Hilfe der Shannonschen Interpolationsformel) rekonstruiert werden.

Die Bedingung an \hat{f} bedeutet anschaulich, daß sich f in jeder Richtung nur mit einer vorgegebenen maximalen Rate ändern darf. Bei der Anwendung des Theorems in der Praxis sind mindestens drei Dinge zu beachten:

- 1. In der Kamera sind die Abstände der Fotozellen festgelegt. Enthält das Bild f sehr feine Bildstrukturen, so wird die Nyquistbedingung $(|\xi| < \frac{1}{2\Delta x} \text{ und } |\eta| < \frac{1}{2\Delta y})$ für f nicht erfüllt sein. Dies führt zu sogenannten Aliasfehlern, die sich als Moiré-Effekte äußern können. Dem entgegen wirkt die Verschmierung des Bildes, die durch die Beugung des Lichts an den nur endlich großen Linsen und Blenden erzeugt wird. Man sollte aber damit rechnen, daß die Nyquistbedingung nicht wirklich erfüllt ist.
- 2. Weil die lichtempfindlichen Zellen in der Kamera eine endliche Ausdehnung haben, findet keine wirkliche Abtastung des kontinuierlichen Bildsignals f statt. Es wird vielmehr eine gemittelte Version g von f abgetastet. Infolgedessen kann man das Abtasttheorem nicht auf f anwenden, sondern höchstens auf g. Das hat aber auch einen Vorteil: Durch die Mittelung erfüllt \hat{g} (bei gegebenem Abstand der Mittelpunkte der Fotozellen) eher die Nyquistbedingung als \hat{f} . Somit ist eine Rekonstruktion von g besser möglich.
- 3. Der Definitionsbereich des Kamerabildes ist beschränkt. Weil Shannons Interpolationsformel jedoch eine unendliche Summe über alle Abtastwerte ist (deren Koeffizienten nur linear gegen Null konvergieren), ist unabhängig von 1. und 2. eine exakte Rekonstruktion von f aus den endlich vielen Abtastwerten nicht möglich.

1.3 Quantisierung

Die Helligkeitsempfindung des menschlichen Auges ist annähernd proportional zum Logarithmus der Lichtintensität – Weber-Fechnersche Gesetz. Deshalb kann man bei vielen Kameras ebenfalls nichtlineare Kennlinien des Verstärkers einstellen; üblich ist $U = I^{\gamma}$, wobei U die Ausgangsspannung und I die Lichtintensität ist. Der Exponent γ – auch *Gammafaktor* genannt – kann verändert werden. Für $\gamma = \frac{1}{2}$ beispielsweise werden die dunklen Bereiche aufgehellt, während $\gamma = 1$ eine lineare Verstärkung bewirkt. **Bemerkung.** Manchmal wird auch $\frac{1}{\gamma}$ als Gammafaktor bezeichnet.

Zur Verarbeitung in einem Digitalrechner werden die Spannungswerte, die den Helligkeiten der Pixel entsprechen, mit einem Analog-Digital-Umsetzer – kurz ADU – quantisiert. Ein ADU ist ein Gerät, das ein beschränktes Intervall aus \mathbb{R} in ein endliches Intervall in \mathbb{Z} abbildet, und zwar in aller Regel vermöge einer monoton wachsenden Treppenfunktion mit Stufen gleicher Länge und Höhe.

Dabei geht natürlich Information des ursprünglichen Signals verloren. Oft wird das auch umgekehrt so modelliert, daß man die Quantisierung als Störung des ursprünglichen Signals durch sogenanntes Quantisierungsrauschen darstellt. Durch Vergrößern der Anzahl der Treppen (für denselben Eingangsspannungsbereich) kann man



Abbildung 1.3: Quantisierung durch monoton wachsende Treppenfunktion

das Quantisierungsrauschen aber im Prinzip beliebig klein machen. Bei Grauwertbildern verwendet man selten mehr als 2^{12} Quantisierungsstufen; oft nur 2^8 . Das menschliche Auge hat zwar einen großen Dynamikbereich (weit mehr als jeder Film), kann aber nur etwa $2^6 = 64$ Grauwerte absolut unterscheiden. Bildgebende Verfahren in der Medizin verwenden meist 12 Bit pro Pixel, die als 16 Bit-Integer abgespeichert werden, um den möglichen Dynamikumfang einzufangen.

1.4 Das Histogramm eines Bildes

Als *Histogramm* bezeichnet man die Häufigkeitsverteilung der Bildwerte $h: W \to [0,1]$. Für ein diskretes Bild $f: R \to W$ mit $0 < \#R < \infty$ ist $h(w) := \frac{1}{\#R} \# f^{-1}(w) = \frac{1}{\#R} \# \{(x,y) \in R \mid f(x,y) = w\}$ für $w \in W$. Manchmal läßt man auch den Normierungsfaktor $\frac{1}{\#R}$ weg.

Bemerkung. Sind R und W Gebiete im \mathbb{R}^2 bzw. \mathbb{R} und $f : R \to W$ eine meßbare Abbildung, so existiert das Bildmaß $\lambda_2 \circ f^{-1}$ des Lebesguemaßes λ_2 im \mathbb{R}^2 . Hat nun $\lambda_2 \circ f^{-1}$ eine Dichte $h : W \to R$ bezüglich des Lebesguemaßes λ_1 auf W, so könnte man h als Histogramm von f bezeichnen. Allerdings ist h nur bis auf Nullmengen bestimmt.

1.5 Aufbau eines Bildverarbeitungssystems für Grauwertbilder

Eine Look-Up-Table – kurz LUT – ist ein Speicher aus k-Bit-Zellen, denen Adressen $0, \ldots, N-1$ zugeordnet sind, unter denen sie angesprochen (das heißt, beschrieben oder ausgelesen) werden können. Mit einer Look-Up-Table kann also jede Abbildung

$$\varphi: \{0, \dots, N-1\} \to \{0, \dots, 2^k - 1\}$$

realisiert werden, indem einfach in der *j*-ten Zelle der Wert $\varphi(j)$ gespeichert wird. Softwaremäßig kann man eine Look-Up-Table als Array realisieren. Verarbeitet man im Bildspeicher *n*-Bit-Bilder, so haben die RGB-LUTs die Länge 2^n . Häufig ist auch k = n = 8; dann stellt jede der LUTs eine Abbildung $[0, 255] \rightarrow [0, 255]$ dar.



Abbildung 1.4: Prinzipieller Aufbau eines Bildverarbeitungssystems

Steht in allen drei LUT an einer Stelle j der gleiche Wert, so erzeugt der Monitor an dieser Stelle j einen grauen Fleck (der in Wirklichkeit aus roten, grünen und blauen Punkten gleicher Leuchtkraft besteht). Realisiert jede LUT die identische Abbildung $\varphi(j) = j$, so hat man die übliche Grauwertwiedergabe.

Durch Abändern der LUTs kann man jedem Wert j z.B. eine Farbe zuordnen. Man spricht dann von einer Falschfarbendarstellung. Dadurch kann man feine Grauwertabstufungen deutlich machen oder gewisse Stellen markieren.

Beispiel. Ordnet man z.B. dem Wert 255 rot zu, indem man $\varphi_R(255) := 255$, $\varphi_G(255) := 0$ und $\varphi_B(255) := 0$ setzt, so werden in jedem Grauwertbild alle Stellen mit Grauwert 255 rot markiert.

1.6 Strategien der Bilderkennung

Bilderkennungsaufgaben verlangen typischerweise festzustellen, ob in einem Bild gewisse Strukturen enthalten sind, z.B. ob Ziffern zu sehen sind und welche. Gesucht wird also eine spezielle Abbildung von dem riesigen Raum aller möglichen Grauwertbilder in den meist sehr kleinen, endlichen Raum von gesuchten Strukturen. Für den Menschen ist es oft leicht, für jedes vorgelegte Bild zu entscheiden, welche der gesuchten Strukturen darin zu sehen ist. Trotzdem kann es schwer fallen, die gesuchte Abbildung algorithmisch zu beschreiben. Denn die gesuchten Strukturen sind meist nicht direkt durch die exakte Verteilung von Grauwerten definiert, sondern eher auf abstraktere Weise durch geometrische Eigenschaften der Helligkeitsverteilung. Deshalb versucht man die gewünschte Abbildung durch eine Hintereinanderschaltung von einzelnen, relativ einfachen Verfahrensschritten approximativ zu konstruieren. Etwas abstrakt formuliert ist es Aufgabe der einzelnen Schritte, die riesige Anzahl von möglichen Bildern durch Äquivalenzklassenbildung allmählich auf immer weniger Alternativen einzuengen, bis zum Schluß nur noch eine Alternativentscheidung zwischen den gesuchten Strukturen übrigbleibt. Für die Wahl der einzelnen Schritte gibt es kein Patentrezept. Sie können recht trivial, aber auch kompliziert und aufwendig sein. Das hängt von der jeweiligen Situation ab, z.B. davon, ob in einer konkreten Anwendung die Grauwertbilder nur von einer speziellen, sehr eingeschränkten Art sind. Man denke an Bankformulare, bei denen in vorgegebenen Kästchen Ziffern einzutragen sind; da sind die Ziffern leichter zu finden und zu identifizieren als bei Mautstellen, wo die Nummernschilder schnell fahrender Fahrzeuge gelesen werden sollen.

Wir wollen dies durch einfache Beispiele verdeutlichen. Als Ausgangsbild wählen wir ein Bild, das die Ziffern Null und Eins zeigt; die Ziffern und der Hintergrund haben jeweils einen konstanten Grauwert. Die Ziffern können durch einfache topologische oder geometrische Kriterien (welche?) unterschieden oder erkannt werden.

1.6. STRATEGIEN DER BILDERKENNUNG



In Anwendungen treten oft typische Störungen wie zufälliges Rauschen oder rampenförmigen Helligkeitsanstieg auf. In der linken Spalte ist jeweils ein so verändertes Bild zu sehen, in der mittleren Spalte der Grauwertverlauf längs der mittleren Bildzeile und in der rechten Spalte das Binärbild, welches entsteht, wenn alle Grauwerte kleiner als 128 durch 0 (schwarz) und alle anderen durch 255 (weiß) ersetzt werden (eine sogenannte Schwellwertentscheidung).



Abbildung 1.5: leicht verrauscht

In dem Binärbild sind die Ziffern nicht mehr so leicht durch einfache Kriterien zu erkennen. Man kann versuchen, das Binärbild durch geeignete Verarbeitungsschritte so zu modifizieren, daß wieder einfache Kriterien greifen. Dies funktioniert nicht bei rampenförmigem Helligkeitsanstieg. In diesem Fall muß man die Entscheidungsschwelle geschickter wählen. Und bei stark verrauschten Bildern muß man zusätzlich die Grauwertbilder vor der Binärisierung vorverarbeiten, Wie dafür geeignete Operationen aussehen, wird ein wichtiger Teil der Vorlesung sein.



Abbildung 1.6: rampenförmiger Helligkeitsanstieg



Abbildung 1.7: Stark verrauscht

In den letzten beiden Beispielen unterscheiden sich die beiden Ziffern nicht mehr durch unterschiedliche (mittlere) Grauwertpegel, sondern nur durch die Musterung (die sogenannte Textur). Hier müssen zunächst aus den Grauwerten an jeder Stelle Texturmerkmale berechnet werden, mit denen dann (vielleicht) eine Unterscheidung von Ziffern und Hintergrund und damit eine geeignete Binärisierung möglich wird.



Kapitel 2

Translationsinvariante, lokale Operatoren im Ortsbereich

Zur Erinnerung: \mathbb{B} ist die Menge der partiellen Abbildungen $f: \mathbb{Z}^2 \xrightarrow{p} \mathbb{R}$.

Eine lokale Filteroperation ist eine Abbildung $\varphi : \mathbb{R}^M \to \mathbb{R}$, wobei $M = \llbracket -n_x, n_x \rrbracket \times \llbracket -n_y, n_y \rrbracket \subset \mathbb{Z}^2$. Wir nennen M das Operatorfenster und n_x bzw. n_y die Filterkernweiten in x- bzw. y-Richtung. Jede lokale Filteroperation φ induziert eine Abbildung $T_{\varphi} : \mathbb{B} \to \mathbb{B}$ vermöge folgender Definition:

Für $f: R \to \mathbb{R}$ mit $R \subset \mathbb{Z}^2$ sei $T_{\varphi}f: D \to \mathbb{R}$ definiert durch

$$T_{\varphi}f(x,y) := \varphi\left((f(x+i,y+j))_{(i,j)\in M}\right) = \varphi\left(\begin{array}{ccc}f(x-n_x,y-n_y)&\cdots&f(x+n_x,y-n_y)\\\vdots&\\f(x-n_x,y+n_y)&\cdots&f(x+n_x,y+n_y)\end{array}\right)$$

für $(x,y) \in D$, wobei $D := \{(x,y) \in \mathbb{Z}^2 \mid (x,y) + M \subset R\}.$

Diese Abbildung T_{φ} ist *translationsinvariant*, das heißt, sie kommutiert mit Translationen in dem folgenden Sinn, daß für eine (beliebige) *Translation* $\tau : \mathbb{Z}^2 \to \mathbb{Z}^2$ mit $(x, y) \mapsto (x, y) + (\xi, \eta)$ und $(\xi, \eta) \in \mathbb{Z}^2$ die Gleichung

$$T_{\varphi}(f \circ \tau) = T_{\varphi}f \circ \tau$$

gilt, wobei $f : R \to \mathbb{R}$ ist. Daher nennt man solche Abbildungen T_{φ} , die durch lokale Filteroperationen φ definiert werden, translationsinvariante, lokale Operatoren oder auch ortsinvariante Filter. Wir werden sie oft einfach als Filter bezeichnen.

Durch Einschränkung erhält man aus einem Filter eine Selbstabbildung $\mathbb{B}(\mathbb{Z}^2, \mathbb{R}) \to \mathbb{B}(\mathbb{Z}^2, \mathbb{R})$ der überall definierten, reellwertigen Bilder. Sind jedoch der Definitionsbereich oder der Wertebereich eines Bildes eingeschränkt, so wird das Filterresultat im allgemeinen einen anderen Definitions- und Wertebereich haben, was bei Anwendungen stören kann.

2.1 Vereinbarung über den Definitionsbereich

Hat ein Bild $f : R \to \mathbb{R}$ einen beschränkten Definitionsbereich R, so hat das Filterresultat $T_{\varphi}f$ einen echt kleineren Definitionsbereich, außer wenn beide Filterkernweiten 0 sind. Dies kann bei praktischen Anwendungen erheblich stören, insbesondere wenn mehrere Filterschritte hintereinander ausgeführt werden. Deshalb setzt man meist $T_{\varphi}f$ durch irgendwelche Ad-hoc-Definitionen auf den Definitionsbereich von f fort; für etliche Filteroperationen φ läßt sich obige Definition von $T_{\varphi}f(x, y)$ in naheliegender Weise so modifizieren, daß sie auch für Punkte nahe des Randes von R noch sinnvoll ist. Wir werden dies jeweils im Einzelfall tun und die entstehenden Abbildungen – etwas mißbräuchlich – wieder translationsinvariante, lokale Operatoren oder Filter nennen. Dabei setzen wir im allgemeinen stillschweigend voraus, daß der Definitionsbereich viel größer als das Operatorfenster ist.

2.2 Vereinbarung über den Wertebereich

In Anwendungen ist der Wertebereich W oft eingeschränkt, z.B. W = [0, 255]. Für viele Filteroperationen φ gilt jedoch $\varphi(W^M) \not\subset W$, z.B. ist der arithmetische Mittelwert natürlicher Zahlen nicht notwendig wieder natürlich. Deshalb werden wir für theoretische Überlegungen ganz \mathbb{R} als Wertebereich wählen. Auch in praktischen Implementierungen ist es sinnvoll, für Zwischenergebnisse den Typ *double* zu verwenden. Will man jedoch ein Filterergebnis als Bild auf dem Monitor darstellen, so muß man daraus ein W-wertiges Bild machen. Sofern alle Werte des Bildes nicht negativ sind, kann man ähnlich wie bei der Quantisierung vorgehen, also z.B. die Werte zum jeweils nächstgelegenen Wert in W runden (etwa durch Type-Cast).

Bemerkung. Es gibt Filteroperationen, die sinnvolle negative Werte produzieren, wie z.B. Differenzenoperatoren, die Ableitungen approximieren; dann ist es nicht sinnvoll, alle negativen Werte durch 0 (dem nächstgelegenen Wert in [0, 255]) zu ersetzen.

Vereinbarung. Ein Bild $f : R \to \mathbb{R}$, das positive und negative Werte annehmen kann, transformieren wir stets auf folgende Weise in ein 8-Bit-Bild g:

$$g(x,y) := \begin{cases} 0 & , \text{ wenn } 128 + f(x,y) < 0\\ 255 & , \text{ wenn } 128 + f(x,y) > 255\\ (unsigned \ char)(128 + f(x,y)) & , \text{ sonst} \end{cases}$$

2.3 Punktoperatoren

Ein Punktoperator ist ein translations
invarianter, lokaler Operator T mit $n_x = n_y = 0$; es gibt also eine Abbildung $\varphi: W \to W$, so daß

$$Tf(x,y) = \varphi(f(x,y))$$
 für $(x,y) \in R$.

Für $W = [0, 2^n - 1]$ läßt sich T durch eine Look-Up-Table realisieren. Beispiele. Typische Anwendungen:

1) Binärisierung von Grauwertbildern: Sei $\Theta \in [0, 255]$. Setze $\varphi(t) = 255 \cdot \chi_{[\Theta, 255]}(t)$.

Ein Problem, das sich in der Praxis ergibt, ist, wie man Θ wählen soll, um die interessanten Bildstrukturen vom Hintergrund abzuheben.

- 2) Äquidensitendarstellung: Hervorhebung derjenigen Bildteile, deren Grauwerte in einem gewissen Bereich liegen. Setze $\varphi(t) = 255 \cdot \chi_{\llbracket\Theta_1,\Theta_2\rrbracket}(t)$ oder mit mehreren Niveaus $\varphi(t) = \sum \omega_i \cdot \chi_{\llbracket\Theta_i,\Theta_{i+1}\rrbracket}(t)$. Spezialfall: Graustufenvergröberung: $\varphi(t) = 2^l \cdot \lfloor \frac{t}{2^l} \rfloor$ ergibt 2^{n-l} statt 2^n Graustufen.
- 3) Hintergrundelimination oder Segmentierung: Ist ein Objekt heller als der Hintergrund, so kann man für ein geeignetes Θ mit φ(t) = t · χ_[Θ,255](t) die Hintergrundstruktur zum Verschwinden bringen.
 Auch hier stellt sich das Problem, daß es nicht immer so eine Schwelle gibt, und falls sie existiert, wie man sie findet.
- 4) Kontraständerung: Ein flaues Bild kann manchmal kontrastreicher gemacht werden, indem man die Helligkeitsabstufungen in gewissen Grauwertbereichen vergrößert, also z.B. φ wählt, wie in der Abbildung skizziert.



Abbildung 2.1: Skizze einer kontraständernden Operation

2.4. LINEARE OPERATOREN

Problem in der Praxis: Wie wählt man a und b?

Oft wählt man ein geeignetes nichtlineares φ manuell interaktiv, z.B. mit "xv". Schwieriger ist eine automatische Wahl; hier bedient man sich oft des Histogramms eines Bildes. Und zwar versucht man, ein monoton wachsendes φ so zu wählen, daß das Histogramm von $T_{\varphi}f$ ausgeglichen wird. Damit meint man idealerweise, daß es konstant wird; dies läßt sich für diskretes W allerdings nicht immer erreichen (im Kontinuierlichem sehr wohl), denn $\varphi \circ f$ hat höchstens so viele *Fasern*, das heißt, Urbilder von Punkten, wie f. Das bedeutet, die Partition des Bildes in die Bereiche konstanter Helligkeit wird durch einen Punktoperator höchstens gröber. Ein Punktoperator kann die Anzahl der in einem Bild vorhandenen Grauwerte nicht vergrößern. Er kann nur die Grauwerte verändern und dadurch ihre Anzahl eventuell verkleinern.

Man ist daher schon zufrieden, wenn die tatsächlich vorkommenden Grauwerte des Bildes annähernd gleich häufig vorkommen und einigermaßen gleichmäßig über den möglichen Wertebereich W verstreut sind. Man läßt also explizit zu, daß etliche Grauwerte überhaupt nicht vorkommen, daß also das Histogramm, wenn man es als Balkendiagramm aufträgt, Lücken hat; nur sollten diese Lücken einigermaßen gleichmäßig verteilt sein.

2.4 Lineare Operatoren

Ein translations invarianter, lokaler Operator heißt *linear*, wenn er von einer linearen, lokalen Filter operation $\varphi : \mathbb{R}^M \to \mathbb{R}$ erzeugt wird. φ ist genau dann linear, wenn es eine Matrix $A = (a(i, j))_{(i,j) \in M} \in \mathbb{R}^M$ gibt, so daß

$$\varphi\left((w(i,j))_{(i,j)\in M}\right) = \sum_{(i,j)\in M} a(i,j)w(i,j)$$

gilt und somit für jedes $f: R \to \mathbb{R}$ und für alle (x, y) mit $(x, y) + M \subset R$

$$T_{\varphi}f(x,y) = \sum_{(i,j)\in M} a(i,j)f(x+i,y+j).$$

Die Matrix A heißt Koeffizientenmatrix oder Maske von T_{φ} . Wir setzen $T_A := T_{\varphi}$. Einen linearen, translationsinvarianten, lokalen Operator nennt man auch lineares Filter. Die Komposition $T_A \circ T_B$ zweier linearer Filter ist wieder ein lineares Filter, dessen Maske jedoch i.a. nicht $A \cdot B$ ist.

Ein lineares Filter mit der Maske $C = (c(i,j))_{(i,j) \in [\![-n_x,n_x]\!] \times [\![-n_y,n_y]\!]}$ heißt separabel, wenn es zwei Masken $A = (a(i,j))_{(i,j) \in [\![-n_x,n_x]\!] \times \{0\}}$ und $B = (b(i,j))_{(i,j) \in \{0\} \times [\![-n_y,n_y]\!]}$ gibt, so daß $C = A \cdot B$.

Man beachte, daß die beiden Masken A und B Vektoren sind, genauer A ist ein Spaltenvektor und B ein Zeilenvektor. Für die Elemente von C gilt c(i, j) = a(i, 0) b(0, j). Sind T_A, T_B, T_C die linearen Filter mit den Masken A, B, C, so gilt:

$$T_C = T_A \circ T_B = T_B \circ T_A.$$

Sinn: Rechenersparnis, denn die Berechnung von T_A und T_B erfordert pro Pixel nur $2n_x + 2n_y + 2$ Multiplikationen und Additionen statt $(2n_x + 1) \cdot (2n_y + 1)$ Operationen bei direkter Berechnung von T_C .

2.4.1 Lineare Glättungsfilter

Ein lineares Filter mit Maske A heißt lineares Glättungsfilter, wenn $a(i,j) \ge 0$ für jedes $(i,j) \in M$ und $\sum_{(i,j)\in M} a(i,j) = 1$. Ein lineares Glättungsfilter bildet in jedem Bildpunkt einen gewichteten linearen Mittelwert über die Grauwerte im Operatorfenster.

Definition von Glättungsfiltern in der Nähe des Bildrandes Für $f : R \to \mathbb{R}$ und $(x, y) \in R$ setze

$$T_A f(x, y) = \frac{1}{\gamma(x, y)} \sum_{(i, j) \in M: (x+i, y+j) \in R} a(i, j) \cdot f(x+i, y+j), \qquad (*)$$

wobei $\gamma(x,y) = \sum_{(i,j) \in M: (x+i,y+j) \in R} a(i,j)$, falls $\gamma(x,y) > 0$, sonst $T_A f(x,y) = 0$.

Man kann auch mit entsprechendem $\gamma(x, y)$ schreiben:

$$T_A f(x,y) = \frac{1}{\gamma(x,y)} \sum_{(u,v) \in R \cap ((x,y)+M)} a(u-x,v-y) \cdot f(u,v).$$
(**)

Für die Pixel (x, y) mit $(x, y) + M \subset R$ ergibt sich wegen $\gamma(x, y) = 1$ die obige allgemeinere Definition von T_A . **Bemerkung.** Verwendet man (*) oder (**) als Definition von T_A , so ist es nicht nötig, daß sich die Filterkoeffizienten zu Eins summieren, weil die Normierung auf das Gesamtgewicht Eins an jeder Stelle (x, y) sowieso durchgeführt wird. Es genügt, daß die Koeffizienten positiv sind. (Fordert man nur nichtnegativ, so könnte an manchen Stellen $\gamma(x, y) = 0$ sein.)

2.4.1.1 Rechteckfilter, Boxfilter

Alle Koeffizienten sind gleich, also $a(i, j) = \frac{1}{(2n_x+1)\cdot(2n_y+1)}$. Ein Boxfilter berechnet an jeder Stelle das arithmetische Mittel über die Grauwerte im jeweiligen Operatorfenster. Man würde erwarten, daß es Bildstrukturen um so stärker verschmiert, je feiner sie sind. Das stimmt aber nicht immer. Man betrachte z.B. ein Bild mit vertikalen Streifen unterschiedlicher Breite:

Ein tieferes Verständnis für dieses Phänomen erhalten wir später mit Hilfe der Fouriertransformation. **Bemerkung.** Rechteckfilter sind separabel.

2.4.1.2 Gaußfilter

Gegeben seien $\sigma_x, \sigma_y \in \mathbb{R}^+$ (meist ≥ 0.5). Wähle $n_x, n_y \in \mathbb{N}$ mit $n_x \geq 2\sigma_x$ und $n_y \geq 2\sigma_y$. Für $(k, l) \in M$ sei

$$\tilde{a}(k,l) = \exp\left(-\frac{k^2}{2\sigma_x^2} - \frac{l^2}{2\sigma_y^2}\right), \ \alpha = \sum_{(k,l)\in M} \tilde{a}(k,l).$$

Das Filter mit der Maske

$$A = \frac{1}{\alpha} \left(\tilde{a}(k,l) \right)_{(k,l) \in M}$$

berechnet an jeder Stelle einen abgeschnittenen Gauß-gewichteten Mittelwert; man nennt es ein *Gaußfilter* mit den *Varianzen* σ_x^2 und σ_y^2 . Ist $\sigma_x = \sigma_y$, so heißt es *isotrop*.

Das Operatorfenster M wird so gewählt, daß die Koeffizienten am Rand genügend klein sind. Meist wählt man $2\sigma_x \le n_x \le 3\sigma_x$ und n_y analog.

Bemerkung. Gaußfilter sind separabel, denn $a(k,l) = \frac{1}{\sqrt{\alpha}} \exp(-k^2/2\sigma_x^2) \cdot \frac{1}{\sqrt{\alpha}} \exp(-l^2/2\sigma_y^2)$.

2.4.1.3 Binomialfilter

 Sei

$$a(k,l) = 2^{-2n_x} \cdot 2^{-2n_y} \binom{2n_x}{n_x + k} \binom{2n_y}{n_y + l} \text{ für } (k,l) \in M.$$

Es gilt: $\sum_{k=-n_x}^{n_x} \binom{2n_x}{n_x+k} = \sum_{k=0}^{2n_x} \binom{2n_x}{k} = 2^{2n_x}$ und entsprechend für n_y , also gilt $\sum a(k,l) = 1$. Die Koeffizienten können leicht mit dem Pascalschen Dreieck berechnet werden. **Bemerkung.** Binomialfilter sind separabel.

12

2.4.2 Differenzenfilter

Als *Differenzenfilter* bezeichnen wir lineare Filter, deren Koeffizienten unterschiedliche Vorzeichen haben. Sie werden meist eingesetzt, um lokale Änderungen im Bild deutlicher hervorzuheben oder zu markieren, z.B. Linien und Kanten. Sie können oft als diskrete Analoga zu linearen Differentialoperatoren verstanden werden. Vereinbarung. Die Koeffizienten der Filtermaske eines Differenzenfilters sollen sich zu Null summieren. Das bewirkt, daß ein Differenzenfilter in größeren Bereichen mit konstantem Grauwert den Ausgangswert Null liefert. Bemerkung. Für Differenzenfilter gibt es meist keine so naheliegende Fortsetzung auf Pixel in der Nähe des Randes des Definitionsbereichs des Bildes wie bei Glättungsfiltern. Lappen Bereiche der Filtermaske, die nicht verschwindende Koeffizienten enthalten, über den Definitionsbereich hinaus, so ist es oft am besten den Ausgangswert auf Null zu setzen. Anderenfalls können in Randnähe unerwünschte Artefakte entstehen, die dann z.B. fälschlicherweise als Bildstrukturen interpretiert werden.

2.4.2.1 Ableitungen erster Ordnung, richtungsabhängige Kantenoperatoren

Übliche diskrete Approximationen von $\frac{\partial f}{\partial x}(x, y)$	Filtermasken	n_x	n_y
f(x,y) - f(x-1,y)	(-1, 1, 0)	1	0
f(x+1,y) - f(x,y)	(0, -1, 1)	1	0
$\frac{1}{2}(f(x+1,y) - f(x-1,y))$	$(-\frac{1}{2}, 0, \frac{1}{2})$	1	0

Analog für $\frac{\partial f}{\partial y}$. Die Filter sind extrem *rauschempfindlich*, das heißt, sie liefern Ausgangswerte ungleich Null, selbst wenn nur kleine Grauwertfluktuationen vorhanden sind, die für den menschlichen Betrachter keine Bildinformation tragen. Um die Rauschempfindlichkeit zu reduzieren, kann man diesen Filtern Glättungsfilter vorschalten, die eine Glättung in der zur Differentiationsrichtung orthogonalen Richtung bewirken. **Beispiele.**

Glättungsfiltermaske Differenzenfiltermaske Maske der Hintereinanderausführung

Box filter	$\frac{1}{3} \left(\begin{array}{c} 1\\ 1\\ 1 \end{array} \right)$	$\begin{pmatrix} -1 & 1 & 0 \end{pmatrix}$	$\frac{1}{3} \left(\begin{array}{rrr} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{array} \right)$
Box filter	$\frac{1}{3}\left(\begin{array}{ccc}1&1&1\end{array}\right)$	$\left(\begin{array}{c} -1\\ 1\\ 0\end{array}\right)$	$\frac{1}{3} \left(\begin{array}{rrr} -1 & -1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{array} \right)$
Binomial filter	$\frac{1}{4} \left(\begin{array}{c} 1\\2\\1 \end{array} \right)$	$\frac{1}{2}\left(\begin{array}{ccc} -1 & 0 & 1 \end{array}\right)$	$\frac{1}{8} \left(\begin{array}{rrr} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{array} \right)$
Binomial filter	$\frac{1}{4}\left(\begin{array}{ccc}1&2&1\end{array}\right)$	$\frac{1}{2}\left(\begin{array}{c} -1\\ 0\\ 1 \end{array} ight)$	$\frac{1}{8} \left(\begin{array}{rrr} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{array} \right)$

Sobel-Operatoren:

$$S_{-} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \ S_{|} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}, \ S_{/} = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix}, \ S_{\backslash} = \begin{pmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

DOB-Filter (difference of boxes): Hintereinanderausführung eines Box- und Differenzenfilters, z.B.

$$\left(\begin{array}{rrrr}1 & 1 & 1\\1 & 1 & 1\\1 & 1 & 1\end{array}\right) \text{ und } \left(\begin{array}{rrrr}-1 & 0 & 1\\-1 & 0 & 1\end{array}\right) \text{ ergibt } \left(\begin{array}{rrrr}-1 & -1 & 0 & 1 & 1\\-1 & -1 & 0 & 1 & 1\\-1 & -1 & 0 & 1 & 1\end{array}\right).$$

2.4.2.2 Ableitungen zweiter Ordnung, isotrope Kantenoperatoren

Es gibt keinen linearen Differentialoperator erster Ordnung, der unter Drehungen invariant ist. Der Laplace-Operator (zweiter Ordnung) $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial u^2}$ dagegen ist isotrop. Eine übliche Approximation von

$$\frac{\partial^2 f}{\partial x^2}(x,y) \text{ ist } (f(x+1,y) - f(x,y)) - (f(x,y) - f(x-1,y)) = f(x+1,y) - 2f(x,y) + f(x-1,y).$$

Man wählt also die Filtermaske $(1 - 2 \ 1)$. Entsprechend wählt man für $\frac{\partial^2 f}{\partial y^2}(x, y)$ die Filtermaske $(1 - 2 \ 1)^T$. Den Laplace-Operator approximiert man folglich durch die Maske

$$\left(\begin{array}{rrrr} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{array}\right).$$

Leider ist diese Maske nicht besonders drehungsinvariant. Man hat daher versucht, Filter zu konstruieren, die ebenfalls den Laplace-Operator approximieren, deren Masken aber etwas besser isotrop sind. Eine Art solcher Filter hat die Gestalt: $-4 \cdot \text{Id} + 4 \cdot \text{Binomialfilter}$.

Für eine 3×3 -Maske ergibt sich das "diskrete 3×3 -Laplace-Filter"

$$-4\left(\begin{array}{rrr} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right) + \frac{1}{4}\left(\begin{array}{rrr} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{array}\right) = \frac{1}{4}\left(\begin{array}{rrr} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{array}\right).$$

Bemerkung. Während bei Glättungsoperatoren durch einen Normierungsfaktor das Gesamtgewicht auf Eins eingestellt wird, hat ein Faktor bei Differenzenfilter keine so intuitive Bedeutung; außer man will einen bestimmten Differentialoperator approximieren. Meist wählt man ihn aber nur mit dem Ziel, das Resultat geeignet zu verstärken.

Bemerkung. In der Literatur haben die Koeffizienten der Laplace-Filter oft das entgegengesetzte Vorzeichen. Damit wird dann $-\Delta$ approximiert.

2.5 Nichtlineare Gradientenoperatoren

Um in kontinuierlichen, differenzierbaren Bildern Stellen zu finden, bei denen sich die Helligkeit stark ändert, kann man die Länge des Gradienten verwenden. Diskrete Gradientenoperatoren sind diskrete Approximationen der Gradientenlänge. Zur Längenmessung verwendet man meist eine der Normen $\|\cdot\|_2$, $\|\cdot\|_1$, $\|\cdot\|_{\infty}$.

Im Kontinuierlichen gilt: $\|\nabla f\|_1 = \left|\frac{\partial f}{\partial x}\right| + \left|\frac{\partial f}{\partial y}\right|, \quad \|\nabla f\|_{\infty} = \max\left\{\left|\frac{\partial f}{\partial x}\right|, \left|\frac{\partial f}{\partial y}\right|\right\}$ und $\|\nabla f\|_2 = \left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2\right)^{1/2} = \left(\left(\frac{\partial f}{\partial u}\right)^2 + \left(\frac{\partial f}{\partial v}\right)^2\right)^{1/2},$ für jedes kartesische Koordinatensystem (u, v).

Beachte: $\|\nabla f\|_1$ und $\|\nabla f\|_{\infty}$ sind nicht invariant unter Rotationen des Koordinatensystems.

Im Diskreten kann man zusätzlich unter verschiedenen Approximationen der partiellen Ableitungen (eventuell auch in Diagonalrichtung) wählen. Dementsprechend viele Operatoren findet man in der Literatur. **Beispiele.**

a) Roberts-Operatoren: Für jedes $q \in \{1, 2, \infty\}$:

$$Tf(x,y) = \|f(x+1,y) - f(x,y), f(x,y+1) - f(x,y)\|_q \text{ oder}$$

$$Tf(x,y) = \|f(x+1,y+1) - f(x,y), f(x+1,y) - f(x,y+1)\|_q$$

b) Sobel-Operatoren:

Tf(x,y) = Summe (oder Maximum) der Werte $|S_{-}f(x,y)|$, $|S_{|}f(x,y)|$, $|S_{/}f(x,y)|$ und $|S_{\setminus}f(x,y)|$.

c) Prewitt-Operatoren:

Tf(x,y) = Summe (oder Maximum) der Werte von DOB-Filtern in mehrere Richtungen.

2.6 Baryzentrische Kantenoperatoren

Differenzenoperatoren, die Differential
operatoren approximieren, werden oft eingesetzt, um in Bildern Stellen mit deutlichem Helligkeitswechsel, sogenannte Helligkeitskanten, zu finden und ihre Richtung zu bestimmen. Statt die Idee der Differentiation kontinuierlicher Bildsignale ins Diskrete zu übertragen, kann man eine andere Idee zur Kantenfindung einsetzen, die aufgrund ihres integralen Ansatzes sehr einfach ins Diskrete übertragen werden kann, ohne besonders rauschempfindlich zu werden. Man faßt das Bildsignal f in einer kreisförmigen Umgebung U(p) eines Punktes $p \in R$ als eine Massenverteilung auf, von der man den Schwerpunkt s_p berechnet. Der Vektor $s_p - p$ zeogt dann in Richtung einer deutlichen Helligkeitszunahme, ähnlich wie der Gradient. Die Übertragung ins Diskrete ist offensichtlich: Die Umgebung U(p) wird durch eine diskrete Approximation einer Kreisscheibe ersetzt, wobei man für kleine Umgebungen einfach Quadrate nimmt, und die Integrale zur Berechnung des Schwerpunktes werden durch Summen ersetzt:

 $m_p := \sum_{q \in U(p)} f(q)$ ist die Gesamtmasse innerhalb der Umgebung U(p). Der Schwerpunkt der Massenverteilung f in U(p) ist gleich $s_p = \frac{1}{m_p} \sum_{q \in U(p)} f(q)q$. Für den Differenzvektor $s_p - p$ ergibt sich

$$s_p - p = \frac{1}{m_p} \sum_{q \in U(p)} f(q)(q-p)$$

Dieser Vektor zeigt in Richtung einer deutlichen Helligkeitszunahme innerhalb der Umgebung U(p). Seine Länge ist beschränkt, denn s_p liegt ja innerhalb von U(p). Lässt man jedoch einfach die Division durch den Normierungsfaktor m_p weg, so ist die Länge des Vektors proportional zum Kontrast von f innerhalb von U(p). Um dies zu sehen, schreibe man f(q) = a + (f(q) - a), wobei a der arithmetische Mittelwert von f in U(p) sei. Verändert man nun die Helligkeitsabweichungen zum Mittelwert, indem man f durch das Signal $g(q) = a + \alpha(f(q) - a)$ mit $\alpha > 0$ ersetzt, so gilt $\sum g(q)(q-p) = \sum a(q-p) + \alpha \sum f(q)(q-p) - \alpha \sum a(q-p) = 0 + \alpha \sum f(q)(q-p) + 0 = \alpha \sum f(q)(q-p)$, weil $\sum_{q \in U(p)} (q-p) = 0$ wegen der Symmetrie von U(p). Somit ist der Operator S, der jedem Pixel p den Vektor

$$Sf(p) = \sum_{q \in U(p)} f(q)(q-p)$$

zuordnet, geeignet, um Richtung und Kontrast von Helligkeitskanten zu bestimmen. Er ist nicht sehr rauschempfindlich und extrem einfach zu implementieren, insbesondere wenn man für U(p) ein kleines in p zentriertes Quadrat wählt. Um ihn isotrop zu machen, müsste man für U(p) im Kontinuierlichen eine Kreisscheibe wählen und im Diskreten eine Gewichtsfunktion auf dem Quadrat U(p) verwenden, deren Wert für jedes Pixel $q \in U(p)$ dem Flächenanteil von q entspricht, der innerhalb der Kreisscheibe liegt. In der Praxis kann man sich diesen Aufwand meist ohne nennenswerten Verlust ersparen und einfach wie oben angegeben ungewichtet über alle Pixel in U(p) summieren. Dann stimmt der Operator mit zwei DOB-Operatoren in x- bzw. y-Richtung überein. Meist genügt es, für U(p) recht kleine Quadrate mit Seitenlänge 3 oder 5 zu wählen.

2.7 Rangordnungsoperatoren

Sei $M = \llbracket -n_x, n_x \rrbracket \times \llbracket -n_y, n_y \rrbracket$ und $m := \#M = (2n_x + 1)(2n_y + 1)$. Für jedes $w = (w(k, l))_{(k,l) \in M} \in \mathbb{R}^M$ sei $\rho(w) \in \mathbb{R}^m$ ein Tupel, das genau die Werte w(k, l) in monoton aufsteigender Reihenfolge enthält. ρ ist also eine Abbildung $\mathbb{R}^M \to \mathbb{R}^m$.

Ein Filter heißt Rangordnungsoperator, wenn seine Filteroperation φ über ρ faktorisiert werden kann, wenn es also eine Abbildung $\psi : \mathbb{R}^m \to \mathbb{R}$ mit $\varphi = \psi \circ \rho$ gibt. Ist ψ linear, so spricht man von einem *linearen* Rangordnungsoperator, obwohl der Operator nicht linear im Sinn von Abschnitt 2.4 sein muß. **Beispiele.** $\psi : \mathbb{R}^m \to \mathbb{R}$ habe die Gestalt $\psi(z_1, \ldots, z_m) = \sum_{j=1}^m a_j z_j$ mit $a_j \in \mathbb{R}$.

- 1) Minimum perator: $a_1 = 1$, alle anderen $a_j = 0$, somit $\varphi = \min$.
- 2) Maximum perator: $a_m = 1$, alle anderen $a_j = 0$, somit $\varphi = \max$.
- 3) Medianoperator: $a_{(m+1)/2} = 1$, alle anderen $a_j = 0$.

- *k-Mittel*: $a_j = \frac{1}{2k+1}$ für $j \in [\![\frac{m+1}{2} k, \frac{m+1}{2} + k]\!]$, alle anderen $a_j = 0$; φ berechnet also das arithmetische Mittel der mittleren 2k + 1 Werte. (4)
- Kantenoperator: $a_l = -1$, $a_{m-l+1} = 1$ für ein $l \in [1, \frac{m-1}{2}]$, alle anderen $a_j = 0$. 5)
- Mid-Range-Operator: $a_1 = a_m = \frac{1}{2}$, alle anderen $a_j = 0$. 6)

Bemerkung. In manchen Situationen ist es günstig, den Begriff des Rangordnungsoperators etwas allgemeiner zu fassen und eine Abhängigkeit vom Grauwert des zentralen Pixels des Operatorfensters zuzulassen. Beispiel (Lokaladaptiver Minimum-Maximum-Operator).

Für $w = (w(k,l)) \in \mathbb{R}^M$ und $\rho(w) = z = (z_1, \dots, z_m)$ sei $\varphi(w) := \begin{cases} z_1 & \text{, falls } w(0,0) - z_1 < z_m - w(0,0) \\ z_m & \text{, sonst} \end{cases}$. Dies bedeutet: $T_{\varphi}f(x,y)$ wird gleich dem Minimum der Werte $(f(x+k,y+l))_{(k,l)\in M}$ gesetzt, wenn f(x,y) with the interval of the set o

näher beim Minimum als beim Maximum liegt und gleich dem Maximum ansonsten.

2.8Morphologische Operatoren

Morphologische Operatoren basieren im wesentlichen auf Minimum- und Maximum-Operationen. Von den in 2.7 eingeführten Rangordnungsoperatoren unterscheiden sie sich vor allem dadurch, daß Umgebungen, über die das Minimum bzw. Maximum gebildet wird, unterschiedliche, nicht unbedingt rechteckförmige Gestalt haben. Es wird also ein sogenanntes Strukturelement \mathcal{U} vorgegeben; das ist einfach eine endliche Teilmenge $\mathcal{U} \subset \mathbb{Z}^2$. Die beiden elementaren morphologischen Operationen sind folgende: Sei $f: R \to \mathbb{R}$.

$$T_{\min,\mathcal{U}}f(x,y) := \min_{(u,v)\in R\cap((x,y)+\mathcal{U})} f(u,v) \quad \text{bzw.} \quad T_{\max,\mathcal{U}}f(x,y) := \max_{(u,v)\in R\cap((x,y)+\mathcal{U})} f(u,v).$$

Morphologische Operatoren werden zunächst hauptsächlich zur Verarbeitung binärer Bilder eingeführt. Sei $f: R \to \{0, 1\}$ ein Binärbild, $B_0 := f^{-1}(0)$ und $B_1 := f^{-1}(1)$. Dann gilt:

$$T_{\min,\mathcal{U}}f(x,y) = \begin{cases} 1 & \text{, wenn } ((x,y) + \mathcal{U}) \subset B_1 \\ 0 & \text{, sonst} \end{cases} \quad \text{Erosion von } B_1 \\ = \begin{cases} 0 & \text{, wenn } ((x,y) + \mathcal{U}) \cap B_0 \neq \emptyset \\ 1 & \text{, sonst} \end{cases} \quad \text{Dilatation von } B_0 \\ T_{\max,\mathcal{U}}f(x,y) = \begin{cases} 0 & \text{, wenn } ((x,y) + \mathcal{U}) \subset B_0 \\ 1 & \text{, sonst} \end{cases} \quad \text{Erosion von } B_0 \\ = \begin{cases} 1 & \text{, wenn } ((x,y) + \mathcal{U}) \cap B_1 \neq \emptyset \\ 0 & \text{, sonst} \end{cases} \quad \text{Dilatation von } B_1 \end{cases}$$

Interpretation: $T_{\min,\mathcal{U}}$ und $T_{\max,\mathcal{U}}$ machen aus einem Binärbild wieder ein Binärbild. Weil ein Binärbild g durch $g^{-1}(0)$ eindeutig bestimmt ist, genügt es anzugeben, wie $B_0 = g^{-1}(0)$ von $T_{\min,\mathcal{U}}$ bzw. $T_{\max,\mathcal{U}}$ verändert wird.

Durch $T_{\min,\mathcal{U}}$ wird B_0 vergrößert, durch $T_{\max,\mathcal{U}}$ verkleinert. (Etwa vergleichbar, wie in der Topologie der Abschluß und das Innere einer Menge). Allerdings wird sozusagen nur eine feste Umgebung der Gestalt \mathcal{U} gewählt. Häufig wird für \mathcal{U} eine kreisförmige Umgebung des Ursprungs gewählt. Dann werden durch $T_{\max,\mathcal{U}}$ von B_0 alle Punkte "weggefressen", die zu nahe am Rand liegen, daher Erosion von B_0 . $T_{\min,\mathcal{U}}$ dagegen fügt zu B_0 sogar noch Punkte aus B_1 hinzu, sofern sie nahe genug am Rand liegen, daher Dilatation von B_0 . $T_{\min, \mathcal{U}}$ füllt kleine Löcher von B_0 auf und verdickt B_0 ; $T_{\max,\mathcal{U}}$ dagegen macht B_0 dünner.

Allgemeine morphologische Operationen erhält man, indem man die elementaren Operatoren $T_{\min,\mathcal{U}}$ und $T_{\max,\mathcal{U}}$ mit eventuell unterschiedlichen Strukturelementen in geeigneter Weise komponiert oder arithmetisch verknüpft.

2.8.1Beispiele

1) $T_{\min,\mathcal{U}} \circ T_{\max,\mathcal{U}}$ Opening, Ouverture von B_0 .

2.8. MORPHOLOGISCHE OPERATOREN



Abbildung 2.2: Von links: Ausgangsbild, Filtermaske \mathcal{U} , Dilatation von B_0 und Erosion von B_0 .

- 2) $T_{\max,\mathcal{U}} \circ T_{\min,\mathcal{U}}$ Closing, Fermeture von B_0 .
- 3) Zylinderhutoperator. Seien $\mathcal{U}_+, \mathcal{U}$ Strukturelemente mit $\mathcal{U}_+ \subset \mathcal{U}$ und h > 0. Setze $\mathcal{U}_- = \mathcal{U} \setminus \mathcal{U}_+$.

$$T_{\mathcal{U}_+,\mathcal{U}_-,h}^{\text{Zyl.}} f(x,y) := \begin{cases} f(x,y) & \text{, wenn } T_{\max,\mathcal{U}_+} f(x,y) - T_{\max,\mathcal{U}_-} f(x,y) > h \\ 0 & \text{, sonst} \end{cases}$$

Veranschaulichung für ein 1-dimensionales Signal:



Abbildung 2.3: Anwendung des Zylinderhutoperators

Setze den Hut von oben auf den Graphen, erhalten bleiben nur die Punkte des Signalgraphen, die den Hut durchstoßen, also die hellen Spitzen, die fein genug sind.

Analog kann man natürlich Zylinderhutoperationen zur Detektion dunkler Spitzen konstruieren.

18 KAPITEL 2. TRANSLATIONSINVARIANTE, LOKALE OPERATOREN IM ORTSBEREICH

Kapitel 3

Bildverbesserung

3.1 Verminderung von Rauschen

Die Erzeugung der Grauwerte der Pixel (im Bildsensor oder durch ein bildgebendes Verfahren) ist oft mit Fehlern behaftet, die mehr oder weniger zufällig erscheinen. Als mathematisches Modell für die Messung wählt man meist stochastische Beschreibungen.

3.1.1 Erinnerung

Eine Zufallsvariable S mit Werten im \mathbb{R}^n ist eine Abbildung $S : \Omega \to \mathbb{R}^n$ von einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, p) mit σ -Algebra \mathcal{A} und Wahrscheinlichkeitsmaß p nach $(\mathbb{R}^n, \mathcal{B}_n)$, die \mathcal{A} - \mathcal{B}_n -meßbar ist, das heißt, für jede Borelsche Menge B ist $S^{-1}(B) \in \mathcal{A}$. $(\mathcal{B}_n$ bezeichne die σ -Algebra der Borelschen Mengen im \mathbb{R}^n).

3.1.2 Additiv verrauschte Bilder

Sei (Ω, \mathcal{A}, p) ein Wahrscheinlichkeitsraum und für jedes Pixel $(x, y) \in R$ sei $S_{(x,y)} : \Omega \to \mathbb{R}$ eine Zufallsvariable. Weiter sei $f_0 : R \to \mathbb{R}$ ein Bildsignal. Dann heißt die Familie

$$f_0(x,y) + S_{(x,y)}, \ (x,y) \in R,$$

von Zufallsvariablen ein stochastisches Bildsignal, das aus f_0 durch additives Rauschen hervorgegangen ist.

Das mathematische Modell für eine Messung eines solchen stochastischen Bildsignals ist die Durchführung eines Zufallsexperiments, das heißt, es wird aus Ω gemäß der Wahrscheinlichkeitsverteilung p eine Stichprobe $\omega \in \Omega$ gezogen. Das gemessene Bildsignal $f: \mathbb{R} \to \mathbb{R}$ hat dann die Grauwerte $f(x, y) = f_0(x, y) + S_{(x,y)}(\omega)$.

Interpretation: Es gibt ein ideales, exaktes Bild f_0 , das pixelweise durch zufällige Störterme additiv verändert wird. Durch jede Messung erhält man eine gemeinsame Realisierung der zufälligen Störterme. Die Zufallsvariablen $S_{(x,y)}$ sind auf ein und demselben Wahrscheinlichkeitsraum Ω definiert.

3.1.3 Multiplikativ verrauschte Bilder

Die Definition ist völlig analog zu 3.1.2; man ersetzt lediglich "+" durch "·": Ein stochastisches Bildsignal, das aus f_0 durch multiplikatives Rauschen hervorgeht, hat die Gestalt $f_0(x, y) \cdot S_{(x,y)}$, $(x, y) \in R$.

Durch Logarithmieren erhält man daraus ein stochastisches Signal, das aus $\log f_0$ durch additives Rauschen hervorgeht. Dadurch erhält man zumindest theoretisch eine Zurückführung auf 3.1.2.(Homomorphe Filter) **Bemerkung.** Wir werden im folgenden nur additives Rauschen betrachten.

3.1.4 Erinnerung

Oft gibt man bei einer Zufallsvariablen $S : \Omega \to \mathbb{R}$ den zugrundeliegenden Wahrscheinlichkeitsraum Ω und sein Wahrscheinlichkeitsmaß p gar nicht explizit an, sondern nur das Bildmaß $\mu = p \circ S^{-1}$, die sogenannte *Verteilung* von S, weil dadurch schon alle interessanten Eigenschaften von S ausgedrückt werden: $\mu(B)$ ist die Wahrscheinlichkeit dafür, daß S einen Wert in $B \in \mathcal{B}$ annimmt.

So gilt z.B.

$$\mathcal{E}(S) = \int_{\Omega} S(\omega) \, dp(\omega) = \int_{\mathbb{R}} x \, d\mu(x) \quad \text{Erwartungwert von } S \text{ und}$$
$$\operatorname{var}(S) = \mathcal{E}\left((S - \mathcal{E}(S))^2\right) = \int_{\mathbb{R}} (t - \mathcal{E}(S))^2 \, d\mu(t) \quad \text{Varianz von } S.$$

Nun wird ein stochastisches Bild durch eine ganze Familie $S_{(x,y)}$ von Zufallsvariablen beschrieben und eine Messung durch eine gemeinsame Realisierung der $S_{(x,y)}$, das heißt, man wählt ein $\omega \in \Omega$, in dem alle $S_{(x,y)}$ ausgewertet werden. Die Familie der gemessenen Störterme $S_{(x,y)}(\omega)$, $(x, y) \in R$, ist durch dieses ω miteinander gekoppelt; die Verteilung der Vektoren $(S_{(x,y)}(\omega) : (x, y) \in R)$ im \mathbb{R}^R (für beliebige Stichproben $\omega \in \Omega$ gemäß p) kann man im allgemeinen nicht aus den Verteilungen der einzelnen Zufallsvariablen $S_{(x,y)}$ herleiten. Es können Abhängigkeiten zwischen ihnen bestehen.

Um sie zu berücksichtigen, muß man die gemeinsame Verteilung der $S_{(x,y)}$ verwenden; dazu definiert man sich die Abbildung $S: \Omega \to \mathbb{R}^R$ durch

$$S(\omega) := \left(S_{(x,y)}(\omega) : (x,y) \in R \right).$$

Sie ist eine Zufallsvariable, und ihre Verteilung, also das Bildmaß $p \circ S^{-1}$ heißt die gemeinsame Verteilung ihrer Komponenten $S_{(x,y)}$.

Die Familie $(S_{(x,y)})_{(x,y)\in R}$ heißt *unabhängig*, wenn die gemeinsame Verteilung der $S_{(x,y)}$ das Produktmaß der Verteilungen der einzelnen $S_{(x,y)}$ ist. (Das heißt, in einem stochastischen Sinn sind sie eben doch nicht gekoppelt.) Für unabhängige Zufallsvariablen werden theoretische Untersuchungen meist viel einfacher. So gilt z.B. **3.1.5 Lemma.** Sind $S_1 : \Omega \to \mathbb{R}$ und $S_2 : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen, so gilt:

$$\mathcal{E}(S_1 \cdot S_2) = \mathcal{E}(S_1) \cdot \mathcal{E}(S_2).$$

Beweis. Seien $\mu_S, \mu_{S_1}, \mu_{S_2}$ die Verteilungen von $S = (S_1, S_2), S_1$ bzw. S_2 und $\varphi : \mathbb{R}^2 \to \mathbb{R}, (t_1, t_2) \mapsto t_1 \cdot t_2$. Dann gilt:

$$\begin{aligned} \mathcal{E}(S_1 \cdot S_2) &= \mathcal{E}(\varphi(S)) \\ &= \int_{\mathbb{R}^2} \varphi(t_1, t_2) \, d\mu_S(t_1, t_2) \\ &= \int_{\mathbb{R}^2} t_1 \cdot t_2 \, d\mu_{S_1} \otimes \, d\mu_{S_2}(t_1, t_2) \\ \overset{\text{Fubini}}{=} \int_{\mathbb{R}} t_1 \, d\mu_{S_1}(t_1) \cdot \int_{\mathbb{R}} t_2 \, d\mu_{S_2}(t_2) \\ &= \mathcal{E}(S_1) \cdot \mathcal{E}(S_2). \end{aligned}$$

3.1.6 Lemma. Sind $S_1, \ldots, S_k : \Omega \to \mathbb{R}$ paarweise unabhängige Zufallsvariablen, so gilt:

$$\operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k}S_{j}\right) = \frac{1}{k^{2}}\sum_{j=1}^{k}\operatorname{var}(S_{j}).$$

Beweis. Die Zufallsvariablen $S'_j := S_j - \mathcal{E}(S_j)$ sind ebenfalls paarweise unabhängig, und es gilt $\mathcal{E}(S'_j) = 0$. Dann gilt:

$$\operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k}S_{j}\right) = \mathcal{E}\left(\left(\frac{1}{k}\sum_{j=1}^{k}S_{j}-\mathcal{E}\left(\frac{1}{k}\sum_{j=1}^{k}S_{j}\right)\right)^{2}\right)$$
$$= \mathcal{E}\left(\left(\frac{1}{k}\sum_{j=1}^{k}\left(S_{j}-\mathcal{E}(S_{j})\right)\right)^{2}\right)$$
$$= \frac{1}{k^{2}}\mathcal{E}\left(\left(\sum_{j=1}^{k}S_{j}'\right)^{2}\right)$$
$$= \frac{1}{k^{2}}\sum_{i=1}^{k}\sum_{j=1}^{k}\mathcal{E}(S_{i}' \cdot S_{j}')$$
$$= \frac{1}{k^{2}}\sum_{i=1}^{k}\sum_{j\neq i}\mathcal{E}(S_{i}') \cdot \mathcal{E}(S_{j}') + \frac{1}{k^{2}}\sum_{j=1}^{k}\mathcal{E}\left((S_{j}')^{2}\right)$$
$$= \frac{1}{k^{2}}\sum_{j=1}^{k}\operatorname{var}(S_{j}).$$

3.1.7 Korollar. Sind $S_1, \ldots, S_k : \Omega \to \mathbb{R}$ paarweise unabhängige Zufallsvariablen und identisch verteilt mit Varianz V_0 , so gilt:

$$\operatorname{var}\left(\frac{1}{k}\sum_{j=1}^{k}S_{j}\right) = \frac{1}{k}\cdot V_{0}.$$

3.1.8 Weißes Rauschen

Eine Familie $(S_{(x,y)})_{(x,y)\in R}$ von Zufallsvariablen nennen wir *weißes Rauschen*, wenn sie unabhängig ist, und wenn die $S_{(x,y)}$ identisch verteilt sind und Erwartungswert Null haben.

Häufig werden additive Störungen als weißes Rauschen modelliert. Dadurch werden theoretische Überlegungen wesentlich erleichtert. Allerdings wird damit nicht jede Anwendungssituation korrekt beschrieben. Denn es gibt durchaus Fälle, in denen die feinen Grauwertfluktuationen, die als Störungen empfunden werden, durch physikalische Prozesse erzeugt werden, die bewirken, daß sie in benachbarten Pixeln durchaus korrelliert und nicht unabhängig sind (z.B. Speckle-Rauschen, Aliaseffekte durch Unterabtastung).

Die einzelnen Variablen $S_{(x,y)}$ von weißem Rauschen haben die gleiche Verteilung. Welche Verteilung einer Anwendung angemessen ist, muß im Einzelfall entschieden werden. Vielfach wählt man eine Gaußverteilung mit der Dichte

$$\frac{1}{\sqrt{2\pi\sigma}}e^{-t^2/2\sigma^2}.\qquad (\sigma^2 \text{ ist die Varianz.})$$

Dafür gibt es zwei Gründe:

- 1) Die theoretische Analyse wird wesentlich vereinfacht.
- 2) Die in praktischen Anwendungen auftretenden Verteilungen haben oft glockenförmige Dichten. Das ist kein Zufall, denn die Störungen entstehen oft durch die Überlagerungen vieler unabhängiger, physikalischer Ereignisse, so daß nach dem zentralen Grenzwertsatz ihre Verteilung mit zunehmender Zahl der Elementarereignisse gegen eine Gaußverteilung strebt.

Wir sprechen vom Gaußschen weißen Rauschen, wenn die einzelnen Variablen Gaußverteilt sind.

3.1.9Verminderung der Rauschvarianz durch zeitliche Mittelung

In manchen Anwendungen kann man von einem stochastischen Bildsignal $f(x,y) = f_0(x,y) + S_{(x,y)}$ mehrere unabhängige Realisierungen f_1, \ldots, f_n erhalten, z.B. wenn man von einer statischen Szene mit einer Kamera nacheinander mehrere Aufnahmen macht. Jede Aufnahme f_j entspricht der Wahl eines Elements ω_j aus dem Wahrscheinlichkeitsraum (Ω, p) , so daß $f_j(x, y) = f_0(x, y) + S_{(x,y)}(\omega_j)$ für jedes Pixel $(x, y) \in R$. Die Störgröße $s_j(x,y) := S_{(x,y)}(\omega_j)$ ist eine Funktion $R \to \mathbb{R}$. Für jedes Pixel (x,y) bildet man nun den arithmetischen Mittelwert:

$$\frac{1}{n}\sum_{j=1}^{n}f_{j}(x,y) = \frac{1}{n}\sum_{j=1}^{n}f_{0}(x,y) + \frac{1}{n}\sum_{j=1}^{n}s_{j}(x,y) = f_{0}(x,y) + \frac{1}{n}\sum_{j=1}^{n}s_{j}(x,y).$$

Sind die Aufnahmen, das heißt, die Wahl der ω_i , unabhängig voneinander, so gilt nach dem starken Gesetz der großen Zahlen

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} s_j(x, y) = \mathcal{E}(S_{(x,y)}) \quad p\text{-fast "uberall.}$$

(Exakt: Die Menge der Folgen $(\omega_j)_{j\in\mathbb{N}}$, für die keine Konvergenz vorliegt, ist eine Nullmenge in $\Omega^{\mathbb{N}}$ bezüglich des Produktmaßes $\otimes_{j \in \mathbb{N}} p.$)

Haben die Störterme $S_{(x,y)}$ den Erwartungswert 0, so gilt mit Wahrscheinlichkeit 1, daß

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^n s_j(x, y) = 0.$$

Die Störungen lassen sich also pixelweise mit Wahrscheinlichkeit 1 beliebig klein machen; man muß nur genügend viele unabhängige Realisierungen des Bildsignals zur Verfügung haben. Die Störvariablen in unterschiedlichen Pixeln müssen nicht unabhängig sein!

Eine zeitliche pixelweise Mittelung findet auch im menschlichem Auge statt. Es ist aber nicht sicher, ob sie linear ist.

In vielen Anwendungen liegt nur ein einziges Bild vor, also nur eine einzige Realisierung der Störvariablen. Dann muß man zu anderen Methoden greifen.

Verminderung der Rauschvarianz durch lokale Mittelwertbildung 3.1.10

Das Bildsignal $f: R \to \mathbb{R}$ sei eine Realisierung eines stochastischen Bildsingals $f_0 + S$, das aus $f_0: R \to \mathbb{R}$ durch Addition von weißem Rauschen $S = (S_{(x,y)})_{(x,y)\in R}$ hervorgegangen ist. Es gibt also ein ω mit f(x,y) = 0 $f_0(x,y) + S_{(x,y)}(\omega)$ für alle $(x,y) \in R$. Die Varianz des Rauschens sei V_0 .

Wir wenden nun auf f ein Rechteckfilter T mit den Kernweiten n_x und n_y an. Für das entstehende Bild T f in Stellen (x, y), die weit genug vom Rand R entfernt sind, gilt:

$$Tf(x,y) = \frac{1}{m} \sum_{(i,j) \in M} f_0(x+i,y+j) + \frac{1}{m} \sum_{(i,j) \in M} S_{(x+i,y+j)}(\omega) = Tf_0(x,y) + g(x,y),$$

wobei g(x, y) eine Realisierung der Zufallsvariablen $Z_{(x,y)} := \frac{1}{m} \sum_{(i,j) \in M} S_{(x+i,y+j)}$ ist. Bemerkung. Was kann man an Tf über f_0 ablesen? Im allgemeinen nichts! Denn, ist das Rauschen z.B. Gaußverteilt, so kann $S_{(x,y)}$ jeden Wert annehmen, und wegen der Unabhängigkeit kann man jede beliebige reelle Funktion als $(x, y) \mapsto S_{(x,y)}(\omega)$ mit geeignetem ω darstellen.

Man kann nur Aussagen stochastischer Natur erwarten.

Nach 3.1.7 gilt $\operatorname{var}(Z_{(x,y)}) = \frac{1}{m}V_0 = \frac{1}{m}\operatorname{var}(S_{(x,y)})$. Die Varianz der Störterme $Z_{(x,y)}$ in dem stochastischen Signal $Tf_0 + Z$ ist also nur $\frac{1}{m}$ der Varianz des Rauschen S. (Allerdings sind die $Z_{(x,y)}$ nicht mehr unabhängig.)

3.2. HERVORHEBUNG VON HELLIGKEITSKANTEN

Wie kann man das ausnutzen? Man verwendet die

Tschebyscheffsche Ungleichung. Sei (Ω, p) ein Wahrscheinlichkeitsraum und $Z : \Omega \to \mathbb{R}$ eine integrierbare Zufallsvariable. Für jedes $\varepsilon > 0$ gilt

$$p\left(\left\{\omega \in \Omega : |Z(\omega) - \mathcal{E}(Z)| \ge \varepsilon\right\}\right) \le \frac{\operatorname{var}(Z)}{\varepsilon^2}$$

Beweis.

$$\operatorname{var}(Z) = \int_{\Omega} (Z - \mathcal{E}(Z))^2 \, dp \ge \int_{\{\omega \in \Omega : |Z(\omega) - \mathcal{E}(Z)| \ge \varepsilon\}} (Z - \mathcal{E}(Z))^2 \, dp \ge \varepsilon^2 \, p(\{\omega \in \Omega : |Z(\omega) - \mathcal{E}(Z)| \ge \varepsilon\}).$$

Weil $\mathcal{E}(S_{(x,y)}) = 0$ für alle (x, y), gilt auch $\mathcal{E}(Z_{(x,y)}) = 0$. Damit folgt, daß für jedes $\varepsilon > 0$ die Wahrscheinlichkeit für $|Z_{(x,y)}| \ge \varepsilon$ höchstens ein *m*-tel so hoch ist wie die für $|S_{(x,y)}| \ge \varepsilon$.

Das Risiko, daß bei der Bildaufnahme der Störterm des geglätteten Bildes in (x, y) größer als ε ist, ist also *m*-mal geringer als im Ausgangsbild. Durch Vergrößern von *m*, also Vergrößern des Operatorfensters, kann man das Risiko großer Störterme im Prinzip beliebig klein machen. Dem sind in der Praxis durch die Größe von *R* Grenzen gesetzt.

Beobachtung: Für das gefilterte stochastische Bild $Tf_0 + Z$ gilt an jeder Stelle (x, y)

$$\mathcal{E}(Tf_0(x,y) + Z_{(x,y)}) = Tf_0(x,y),$$

während für das stochastische Ausgangsbild $f_0 + S$ gilt

$$\mathcal{E}(f_0(x,y) + S_{(x,y)}) = f_0(x,y).$$

Problem: Durch die Mittelwertbildung kann man zwar die Wahrscheinlichkeit für große Störungen verringern, aber der Erwartungswert ist nicht mehr der ideale Signalwert $f_0(x, y)$, sondern ein Mittelwert von f_0 über eine Umgebung von (x, y).

 $f_0(x, y)$ und $Tf_0(x, y)$ weichen meist stark voneinander ab, wenn f_0 sich innerhalb des Operatorfensters bei (x, y) stark ändert. Eine Vergrößerung des Operatorfensters steigert dieses Risiko. Man steckt also im folgenden

Dilemma: Je stärker ein Rechteckfilter die Störterme reduziert, desto stärker verschmiert es auch das unverrauschte ideale Signal.

Dies gilt nicht nur für Rechteckfilter, sondern für alle linearen Glättungsfilter. Man kann lediglich die Verschmierung benachbarter Kanten etwas verringern, wenn man glockenförmig abfallende Gewichte verwendet.

3.1.11 Verminderung von Impulsrauschen

Als *Impulsrauschen (Salt & Pepper Noise)* bezeichnet man Störungen, die nur in relativ wenig, isoliert liegenden Pixeln auftreten, dafür aber oft große Abweichungen von dem Grauwert des idealen, ungestörten Bildsignals bewirken.

Zur Beseitigung von Impulsrauschen braucht man eine Filteroperation, die Mittelwerte bildet, die gegen einzelne "Ausreißer" unempfindlich sind. Lineare Mittelungen sind dafür ungeeignet. Recht gute Resultate erzielt man mit dem Median oder ähnlichen Mittelwertbildungen (Robuste Schätzer in der Statistik), sofern die Zahl der gestörten Pixel nicht zu groß ist.

3.2 Hervorhebung von Helligkeitskanten

Die Stellen eines Bildes, in denen sich die Helligkeit abrupt ändert, werden *Helligkeitskanten* genannt. Entsprechend versteht man unter *Farbkanten* oder *Texturkanten* Stellen, an denen sich die Farbe oder Musterung (Textur) stark ändert. Wir betrachten hier nur Helligkeitskanten und nennen sie schlechthin *Kanten*.

Es gibt keine mathematische Definition des Begriffs Kante, die in allen Situationen den visuellen Eindruck eines menschlichen Betrachters adäquat beschreibt. Dieser Eindruck ist nämlich oft abhängig vom Kontext der umliegenden Bildstrukturen, z.B. spielt es eine Rolle, welche Ausdehnung die Grenze zwischen zwei Grauwertbereichen hat, und ob diese Bereiche homogen oder stark struktruiert (z.B. verrauscht) sind. Einfache Definitionen lokaler Natur versagen da. Stattdessen begnügt man sich meist mit einfachen, aber typischen Modellen für Kanten.

3.2.1 Modelle für geradlinige Kanten

Der Einfachheit halber betrachten wir nur senkrechte Kanten; durch Drehungen der Koordinaten erhält man andere Richtungen.

Zur Modellierung verwendet man Bilder $f : R \to \mathbb{R}$, die in jeder Spalte konstant sind, also die Gestalt $f(x, y) = \varphi(x)$ haben, wobei $\varphi : [0, N_x - 1] \to \mathbb{R}$ der Grauwertverlauf längs der Zeilen ist, durch den man die Gestalt der Kante vorgibt, z.B.



Für theoretische Untersuchungen bevorzugt man oft kontinuierliche Modelle, also $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(x, y) = \varphi(x)$ und $\varphi : \mathbb{R} \to \mathbb{R}$. Dann kann man auch stetig differenzierbare Rampen modellieren.

3.2.2 Andere Kantenmodelle

• Gekrümmte Kanten, also Bilder, in denen homogene Grauwertbereiche längs einer gekrümmten Kurve aneinandergrenzen. Insbesondere auch Ecken.

- Linien
- Verrauschte Kanten: Addition von Rauschen zu einem der obigen Kantenmodelle.

3.2.3 Was heißt Hervorhebung von Kanten?

Die meisten Verfahren beruhen auf folgenden Effekten:

- Die Kanten werden durch sehr helle oder sehr dunkle Grauwerte markiert; die anderen Grauwerte des Bildes werden eliminiert, das heißt, zu einem mittleren Grauwert (128) abgeändert. Die Markierung der Kanten kann darin bestehen, daß die Kantenpunkte selbst hell oder dunkel wiedergegeben werden, oder daß die Kante eingesäumt wird von einem hellen Saum auf der einen Seite und einen dunklen auf der anderen.
- 2. Längs der Kanten wird eine Kontrastüberhöhung bewirkt.



Abbildung 3.1: Grauwertverlauf senkrecht zur Kantenrichtung

3. In den durch Kanten begrenzten Bereichen wird die Grauwertverteilung homogenisiert, möglichst ohne die Kanten zu verschmieren. (Kantenerhaltende Glättung)

3.2. HERVORHEBUNG VON HELLIGKEITSKANTEN

Verfahren, die 1. oder 2. bewirken, basieren auf eher differentiellen Methoden, benutzen also meist Differenzenoperatoren. Bei 3. handelt es sich eher um einen integralen Ansatz. Kombiniert man Verfahren 3. mit Verfahren für 1. oder 2. in geeigneter Weise, so kann man Methoden zur Kantendetektion erhalten, die recht unempfindlich gegenüber Störeinflüssen sind.

3.2.4 Ableitungen erster Ordnung

Differenzenoperatoren, die Richtungsableitungen approximieren (z.B. Sobel- und DOB-Operatoren), können zur Hervorhebung von mehr oder weniger geradlinigen Kanten verwendet werden, wenn die Ableitungsrichtung senkrecht zur Kantenrichtung ist.



Ein richtungsabhängiger Differenzenoperator ist zur Hervorhebung von Ecken nicht so gut geeignet; besser ist es, einen richtungsunabhängigen Operator zu verwenden, der die Gradientenlänge schätzt. Allerdings kann man dann nicht am Vorzeichen die Steigung ablesen.

3.2.5 Ableitungen zweiter Ordnung

Zweite Ableitung senkrecht zur Kantenrichtung



Beachte: Sind auf den Parallelen zur Kantenrichtung die Grauwerte konstant, so berechnet der Laplace-Operator $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ die zweite Ableitung senkrecht zur Kante.

Die Kanten werden also durch helle und dunkle Säume eingesäumt. **Bemerkung.** Je steiler die Kante, desto schmäler sind die Säume.

3.2.6 Hochpaßfilter

Eine alte Technik, die schon früher bei Röntgenbildern verwendet wurde, ist folgende Hochpaßfilterung (der Name wird im 4. Kapitel verständlich werden). Sei T_{glatt} ein lineares Glättungsfilter. Setze

$$Hf(x,y) := f(x,y) - T_{glatt}f(x,y).$$

Es gilt dann: $Hf = f - T_{glatt}f = (Id - T_{glatt})f$. (Vergleiche das Ergebnis mit dem des Laplace-Operators.)



Dieses Filter läßt alle feinen Grauwertfluktuationen durch und umrandet sie mit Hell-Dunkel-Säumen. Oft ist es aber wünschenswert, daß nur kontrastreichere oder größere Bildstrukturen umsäumt werden. Dies kann man bis zu einem gewissen Grad durch eine nichtlineare Modifikation erreichen. Wir wählen als Glättungsfilter ein Gaußfilter $T_{\text{Gauß}}$. Dann gilt für $p = (x, y) \in R$:

$$T_{\text{Gauß}}f(p) = \frac{1}{\alpha_p} \sum_{q \in R \cap (p+M)} f(q) g_{\sigma}(\|q-p\|_2),$$

wobei M das Operatorfenster, $g_{\sigma}(t) = e^{-t^2/2\sigma^2}$ und $\alpha_p = \sum_{q \in R \cap (p+M)} g_{\sigma}(\|q-p\|_2)$ ist. Damit gilt:

$$Hf(p) = f(p) - \frac{1}{\alpha_p} \sum_{q} f(q) g_{\sigma}(||q - p||_2) = \frac{1}{\alpha_p} \sum_{q} (f(p) - f(q)) g_{\sigma}(||q - p||_2).$$

Die nichtlineare Modifikation besteht nun darin, daß alle Summanden weggelassen werden, für die |f(p) - f(q)|unterhalb einer vorgegebenen Schwelle liegt. Wir definieren:

$$NLH_{\tau,\sigma}f(p) = \sum_{q} \psi(f(p) - f(q)) g_{\sigma}(||q - p||_2), \quad \text{wobei} \quad \psi(t) := \begin{cases} 0 & \text{, für } |t| \le \tau \\ t & \text{, sonst} \end{cases}$$

Dies ist ein nichtlineares Filter, das ähnliche Resultate wie H liefert; allerdings reagiert es nicht auf kleine Abweichungen.

3.2.7 Kontrastüberhöhung bei Kanten

T sei ein Filter, das entlang der Kanten Hell-Dunkel-Säume erzeugt, z.B. ein Laplace-Filter oder ein Hochpaßfilter. Für $\lambda > 0$ ist dann $f + \lambda T f$ ein Bild, das sich von f lediglich in der Nähe von Kanten unterscheidet, und zwar ist in Abhängigkeit von λ die Höhe der Kanten vergrößert.

3.3 Kantenerhaltende Glättung

Durch lineare Glättungsverfahren werden Kanten unweigerlich verschmiert. Eine kantenerhaltende Glättung läßt sich höchstens durch nichtlineare Methoden realisieren. Wir stellen hier ein Verfahren vor, das auf nichtlinear modifizierten Gaußfiltern beruht. In der Literatur werden solche Filter auch bilaterale Filter genannt. Anschließend skizzieren wir einige weitere Verfahren, die auf völlig anderen Ansätzen beruhen.

3.3.1 Die nichtlineare Gaußfilterkette

Lineares Gaußfilter

$$LG_{\sigma}f(p) = \frac{1}{\alpha_p} \sum_{q \in R \cap (p+M)} f(q) g_{\sigma}(\|q-p\|_2),$$

wobei $p = (x, y) \in R, q \in R, g_{\sigma}(t) = \exp(-t^2/2\sigma^2)$ und $\alpha_p = \sum_q g_{\sigma}(||q - p||_2)$.

Nichtlineares Gaußfilter

$$NLG_{\sigma,\eta}f(p) = \frac{1}{\alpha_p} \sum_{q \in R \cap (p+M)} f(q) g_{\sigma}(\|q-p\|_2) g_{\eta}(f(q) - f(p)),$$

wobei $\alpha_p = \sum_q g_\sigma(\|q-p\|_2) g_\eta(f(q) - f(p)).$

Bei hohem Signal-Rausch-Verhältnis kann man das Rauschen reduzieren, ohne die Kanten zu verschmieren, indem man ein lineares Glättungsfilter so modifiert, daß in den Mittelwert an einer Stelle p nur die Werte f(q)an Nachbarstellen q eingehen, die von f(p) höchstens um eine vorgegebene Schranke abweichen. Anschaulich bedeutet dies, daß nur die Punkte (q, f(q)) des Signalgraphen gemittelt werden, die sich innerhalb eines in (p, f(p))zentrierten Fensters befinden. In Abbildung 3.3.1 sind zwei solche Datenfenster an den Signalsprungstellen 64 und 320 eingezeichnet. Solche Filteroperationen sind als robuste Schätzer in der Statistik bekannt.

3.3. KANTENERHALTENDE GLÄTTUNG

Weil das Rauschen in Anwendungen oft glockenförmige Verteilungen hat, ersetzen wir die harte Auswahl der zu mittelnden Werte durch eine weichere Entscheidung, indem wir sowohl in Orts- wie in Werterichtung statt durch ein Fenster durch Gauß-förmige Gewichtsverteilungen abschneiden. Dadurch ergibt sich das obige nichtlineare Gaußfilter.



Filterung eines 1-dimensionalen Signals

Die Breite der Gaußglocke g_{η} in Werterichtung sollte einerseits groß sein, damit das Rauschen sicher eingefangen wird, andererseits jedoch klein sein, damit die Kanten nicht verschmiert werden. Diese gegensätzlichen Kriterien lassen sich nur bei hohem Signal-Rausch-Abstand erfüllen, wie z.B. in obiger Abbildung bei den beiden hohen Sprüngen rechts; die niedrigen Sprünge links werden verschmiert. Diesem Dilemma kann man durch eine geeignete Hintereinanderschaltung mehrerer nichtlinearer Gaußfilter entkommen [14].

Die nichtlineare Gaußfilterkette

$$f_{aus} = NLG_{\sigma_k,\eta_k} \circ \ldots \circ NLG_{\sigma_1,\eta_1}(f)$$

Jeder Filterschritt sollte einerseits das Rauschen möglichst stark reduzieren, andererseits sollten die Kanten möglichst wenig verschmiert werden. Dazu müssen die Werte der η_j fallen und die der σ_j wachsen. Man kann zeigen, daß $\eta_{j+1} = \eta_j/2$ und $\sigma_{j+1} = 2\sigma_j$ eine optimale Wahl für geringe Kantenverschmierung und große Rauschreduktion ist [13]. In der Abbildung unten wird dies für ein eindimensionales Signal demonstriert, indem an den Sprungstellen 64 und 320 (harte) Fenster eingezeichnet sind, deren Breite und Höhe den Werten von σ und η der folgenden Filterstufe entspricht. Kanten, die in den ersten Filterschritte verschmiert wurden, können sich in den nachfolgenden wieder aufrichten [15]. In der Praxis sind Filterstufen fast keine Veränderung mehr bewirken.



Mehrstufige Filterung eines 1-dimensionalen Signals

Das Ausgangssignal f_{aus} einer solchen Filterkette ist stückweise nahezu konstant, und die Kanten sind leicht detektierbar. Daher liegt es nahe, einen weiteren modifizierten Filterschritt anzuhängen, bei dem die Filtergewichte wie bisher aus f_{aus} berechnet werden, jedoch nicht die Werte von f_{aus} , sondern die des ursprünglichen Eingangssignals f gemittelt werden.



Glättung durch eine 5-stufige nichtlineare Gaußfilterkette

Separierte nichtlineare Gaußfilter

Nichtlineare Gaußfilter sind nicht separabel; denn die Filtergewichte hängen vom Grauwert des zentralen Pixels ab. Die Hintereinanderausführung zweier nichtlinearer Gaußfilter NLG_1 und NLG_2 , deren Operatorfenster nicht quadratisch, sondern von der Gestalt $\{0\} \times [\![-n, n]\!]$ bzw. $[\![-n, n]\!] \times \{0\}$ sind, liefert also i.a. nicht dasselbe Ergebnis wie ein nichtlineares Gaußfilter NLG mit Operatorfenster $[\![-n, n]\!] \times [\![-n, n]\!]$ und ansonsten gleichen Parametern σ und η . Meist ist der Unterschied aber nicht groß und für die Anwendung unwesentlich. Weil die beiden Filter NLG_1 und NLG_2 rein eindimensional auf den Zeilen bzw. Spalten des Bildes arbeiten, benötigen sie sehr viel weniger Rechenzeit als NLG. Daher sind auch große Filterweiten n wie in den letzten Stufen einer Filterkette problemlos realisierbar.

Die Hintereinanderausführung zweier solcher eindimensionaler, nichtlinearer Gaußfilter, von denen das eine in Zeilenrichtung und das andere in Spaltenrichtung arbeitet, nennen wir ein separiertes, nichtlineares $Gau\beta$ filter.

Bemerkung

Nichtlineare Gaußfilter lassen sich in völlig analoger Weise auch für reelle Signale f definieren, deren Definitionsbereich höherdimensional ist, z.B. für Volumendaten. In den obigen Formeln ging nirgends die Dimension des Definitionsbereichs ein. In höheren Dimensionen ist die Rechenzeitersparnis der separierten Filter gegenüber den nichtseparierten enorm!

Literaturhinweise: [13], [14], [15], [16]

3.3.2 Anisotrope Diffusion

Eine lineare Gaußfilterung mit Varian
z σ^2 kann auch als Diffusionsprozeß (Wärmeleitungsprozeß) verstanden werden, der zur Zeit
 $\frac{1}{2}\sigma^2$ gestoppt wird. Das Ergebnis ist also eine Lösung der linearen partiellen Differential-gleichung

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \operatorname{div} \nabla u \quad \operatorname{mit} \quad u(x, y, 0) = f(x, y).$$

("div" und " ∇ " sollen sich hier stets nur auf die Koordinaten x und y beziehen.)

Ein Gaußfilter ist als Glättungsfilter für einen Kantenfinder insofern ungünstig, als es nicht nur das Rauschen, sondern auch die zu detektierenden Sprünge verschmiert.

3.3. KANTENERHALTENDE GLÄTTUNG

Idee: Modifiziere die Diffusions-Differentialgleichung so, daß die Wärmeleitung von (x, y) aus in Richtung ∇u zum Zeitpunkt t umso kleiner wird, je größer $\nabla u(x, y, t)$ ist. Meist verwendet man eine Differentialgleichung der Gestalt

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(\|\nabla u\|) \cdot \nabla u),$$

wobei g positiv und monoton fallend (und so, daß $g(\|\nabla u\|) \in C^1$) ist, z.B. eine Gaußfunktion $s \mapsto e^{-s^2/2k^2}$; mit k kann die Anisotropie der Wärmeleitung gesteuert werden.

Probleme und Nachteile:

- 1. Man muß eine nichtlineare partielle Differentialgleichung lösen. Das geht nur mit numerischen Methoden und erfordert einigen Aufwand.
- 2. Man muß ∇u berechnen, was im Diskreten nur durch Approximation mit geeigneten Differenzen geht. Ist das Bildsignal f verrauscht, so ist die Schätzung von ∇u für kleine Zeiten t ebenfalls verrauscht. Diese Störung pflanzt sich bei der Lösung der Differentialgleichung mit wachsender Zeit fort.

Literaturhinweis: [7], [9] und [11].

3.3.3 Nichtlineare Gaußfilterketten und anisotrope Diffusion

Nichtlineare Gaußfilter kann man auch als eine Art anisotroper Diffusion interpretieren. Betrachten wir die übliche, lineare Wärmeleitungsgleichung

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \operatorname{div} \nabla u \quad \operatorname{mit} \quad u(x, y, 0) = f(x, y).$$

Ihre Lösung ist darstellbar als Faltungsintegral

$$u(p,t) = \frac{1}{4\pi t} \int_{\mathbb{R}^2} e^{-\frac{\|q-p\|^2}{4t}} f(q) \, dq.$$

Für festes $t = \frac{1}{2}\sigma^2$ ist

$$f \mapsto G_{\sigma_x} f(p) = u(p, \frac{1}{2}\sigma_x^2) = \frac{1}{2\pi\sigma^2} \int_{\mathbb{R}^2} g_{\sigma}(||q-p||) f(q) dq$$

ein lineares Gaußfilter mit Varianz σ_x .

Statt nun die Wärmeleitungsgleichung in einer anisotropen Weise zu modifizieren, kann man auch die obige Integraldarstellung der Lösung in geeigneter, anisotroper Weise modifizieren:

$$G_{\sigma,\eta}f(p) = \frac{1}{N(p,f)} \int_{\mathbb{R}^n} g_{\sigma}(\|q-p\|) g_{\eta}(|f(q)-f(p)|) f(q)$$

mit $N(p,f) = \int_{\mathbb{R}^n} g_{\sigma}(\|q-p\|) g_{\eta}(|f(q)-f(p)|)$

Das ist ein nichtlineares Gaußfilter.

Es ist leichter zu implemetieren als die numerische Lösung einer anisotrop modifizierten, partiellen Differentialgleichung. Eine Kette von nichtlinearen Gaußfiltern

$$NLG_{\sigma_k,\eta_k} \circ \ldots \circ NLG_{\sigma_1,\eta_1}$$

kann man interpretieren als eine Art anisotrope Diffusion, die man zu den Zeitpunkten $t_j = \frac{1}{2} \sigma_j^2$ neu startet und zwar mit dem geänderten Parameter η_{j+1} . Wenn man die weiter oben genannte, optimale Strategie verwendet, bei der nach jedem Filterschritt der Parameter η_j halbiert wird, so stoppt die Diffusion automatisch, wenn η_j kleiner als 1 wird.

3.3.4 Die Membranmethode von Blake und Zisserman

Idee: f sei das Bildsignal. Approximiere graph(f) durch eine dünne Metallmembran ϕ , so daß $d + \lambda E + \alpha P$ minimal wird, wobei $\alpha \ge 0, \lambda \ge 0$ zwei Gewichtsfaktoren sind und

d = Summe der Quadrate der euklidischen Abstände der Punkte $(r, f(r)), r \in R$, im Graph von f zu ϕ ,

E =Energie, die für das Biegen der Membran aufgewendet wird,

P = Energie, die zum Reißen der Membran aufgewendet wird.

Eindimensionales Analogon: Approximiere ein diskretes eindimensionales Signal durch einen Stab, der gebogen und gebrochen werden darf.



Der Term d kann als Summe der Energien von Federn aufgefaßt werden, die zwischen den Datenpunkten und dem Stab (oder der Membran) gespannt sind.

Bemerkung. Im Gegensatz zur Splineapproximation ist erlaubt, daß der Stab oder die Membran an nicht vorher festgelegten Stellen reißen darf. Diese Bruchstellen werden als Kantenpunkte markiert.

Mit den Faktoren kann man einstellen, wie starr (Größe von λ) oder wie reißfest (Größe von α) die Membran sein soll. Sie spielen eine ähnliche Rolle wie die Varianz σ in Cannys Kantenfinder (vgl. 4.3). Variiert man sie, so erhält man ebenfalls eine Art von *Scale-Space-Darstellung* des Signals f, allerdings zweiparametrig. Je größer λ und α sind, desto weniger Details des Bildes bleiben erhalten.

Gegenüber der Skalierung mit Hilfe von Gaußfiltern gibt es einen großen Vorteil: Die Kanten, das heißt, die Bruchstellen der Membran, werden kaum verrückt, wenn die Skalierung sich ändert. Wird also z.B. eine geradlinige Sprungkante in einer Skalierungsstufe detektiert, so bleibt sie auch bei Verkleinerung von λ und α an derselben Stelle.

Obwohl diese Methode von Blake und Zisserman gute Resultate liefert, gibt es auch Probleme und Nachteile:

- 1. In praktischen Implementierungen wird die Membran durch eine diskrete Parametrisierung $\varphi(r), r \in R$, gegeben. Die Biegeenergie E wird im kontinuierlichen Fall durch einen Term mit zweiten Ableitungen modelliert, den man im Diskreten durch Differenzieren approximieren muß. Die Reißenergie P sollte nicht nur von der Anzahl der Pixel abhängen, in denen die Membran reißt; sinnvoller wäre es, die Länge der Kurvenstücke, längs derer die Membran reißt, eingehen zu lassen. Das ist technisch schwierig.
- 2. Das Funktional $d + \lambda E + \alpha P$ ist hochgradig nichtkonvex und hat viele lokale Minima. Es ist deshalb schwierig, ein globales Minimum zu finden. Blake-Zisserman geben einen Algorithmus an, von dem aber meines Erachtens nicht bewiesen ist, daß er stets gegen ein globales Minimum konvergiert. (Gradually Non-Convex-Algorithmus)

Eine andere Möglichkeit zur Minimierung bieten stochastische Algorithmen. Beide Ansätze sind numerisch aufwendig und nicht gut parallelisierbar; sie benötigen viele Iterationsschritte.

3.3.5 Stochastische Relaxation, Simulated Annealing

Ein Bild wird modelliert als Realisierung der Summe zweier Familien $\{F_r + Z_r\}_{r \in R}$, wobei $(F_r)_{r \in R}$ eine Familie von nicht unabhängigen Zufallsvariablen ist, deren Realisierungen gerade die idealen, ungestörten Bilder sind, und $(Z_r)_{r \in R}$ eine Familie von unabhängigen, identisch verteilten, normalverteilten Zufallsvariablen ist, die auch von allen F_r unabhängig sind und deren Realisierungen die Störterme sind. Man nimmt an, daß $(F_r)_{r \in R}$ ein Markovfeld ist, das heißt, es gibt ein Umgebungssystem $(U_r)_{r \in R}$, so daß für jedes $r \in R$ die bedingte Verteilung von F_r unter $(F_s)_{s \neq r}$ gleich der unter $(F_s)_{s \in U_r \setminus \{r\}}$ ist. Dann läßt sich die gemeinsame Verteilung der F_r als sogenannte Gibbs-Verteilung in der Form $p(f) = \alpha e^{-V(f)}$ mit $V(f) = \sum_r V_{U_r}(f)$ darstellen, wobei V_{U_r} nur von den f(s) mit $s \in U_r$ abhängt; α ist ein Normierungsfaktor.
3.4. HERVORHEBUNG VON LINIEN DURCH VERDÜNNUNG

Eine entsprechende Darstellung hat die a-posteriori-Verteilung für jedes gemessene Bild g

$$p(F = f \mid F + Z = g) = \alpha e^{-V(f) - \|f - g\|^2 / 2\sigma^2}, \quad \sigma \text{ Varianz der } Z_r.$$

Als Schätzung wählt man ein f, für das p(F = f | F + Z = g) maximal ist oder äquivalent $V(f) + \frac{||f-g||^2}{2\sigma^2}$ minimal ist. (Blake-Zissermans Funktional ist ein Spezialfall)

Somit ist wieder eine Minimierungsaufgabe zu lösen.

Bemerkung. Unsere Darstellung ist stark vereinfacht. Insbesondere betrachten Gemans nicht nur die Familie (F_r) , deren Realisierungen die ungestörten Bilder sind, sondern noch einen sogenannten *line process L*, der die Existenz von Kanten zwischen den Pixeln beschreibt (analog zum Reißen der Membran von Blake-Zisserman). V hängt auch noch von den Werten von L ab.

Zur Minimierung von $V(f) + \frac{\|f-g\|^2}{2\sigma^2}$ verwenden Gemans ein stochastisches Optimierungsverfahren, nämlich stochastische Relaxation mit Simulated Annealing. (Weitere Stichwörter: Metropolis-Algorithmus, Gibbs-Sampler)

Das Verfahren ist aufwendig, weil meist sehr viele Iterationsschritte nötig sind; jeder einzelne Schritt ist allerdings einfach. Es gibt Ansätze zur Parallelisierung. Wegen des großen Aufwandes wird es beim Maschinellen Sehen nur selten eingesetzt.

Literaturhinweis: [3].

3.4 Hervorhebung von Linien durch Verdünnung

Kantenerhaltende Glättungsverfahren haben oft Schwierigkeiten, dünne Linien zu erhalten, weil sie zu wenig flächenhaft ausgedehnt sind. Insbesondere wenn sie ein wenig verwaschen sind, werden sie oft wie andere feine Strukturen geglättet. Diese Situation tritt z.B. bei Röntgenbildern von kontrastmittelgefüllten, dünnen Adern auf (Angiographiebilder).

Stellt man sich den Graphen der Grauwertfunktion als ein Gebirge vor, so sind dünne helle Linien Grate von Bergrücken und dunkle Linien Talsohlen. Je verwaschener die Bilder sind, desto flacher verlaufen die Flanken der Bergrücken bzw. der Täler. Um nun Linien deutlicher sichtbar zu machen, kann man versuchen, die Flanken zu versteilern. Dazu muss man zunächst erkennen, ob ein Pixel zu einer solchen Flanke gehört oder zu einem Grat bzw. zu einer Talsohle.

Wir betrachten den Fall heller Linien. Dunkle Linien kann man analog behandeln. Ein Punkt auf einem Grat zeichnet sich dadurch aus, dass in Richtung des Grates die Höhe sich kaum ändert, also der Grat nahezu eine Höhenlinie ist, während in zum Grat transversalen Richtungen die Höhe stark abfällt. Für einen Punkt auf einer Bergflanke gilt dagegen, dass auf der einen Seite der durch ihn gehenden Höhenlinie die Höhe stark ansteigt und auf der anderen Seite stark abfällt. Daher wählt man eine kleine Kreisscheibe U mit Mittelpunkt 0 und betrachtet für $u \in U$ die Ausdrücke $\gamma(u) = (f(p+u) - f(p)) \cdot (f(p-u) - f(p))$. Verläuft die Höhenlinie durch p ungefähr in Richtung von u, so wird $\gamma(u)$ klein sein. Ist u jedoch transversal dazu, so ist der Betrag $|\gamma(u)|$ relativ groß, und $\gamma(u)$ selbst ist positiv, wenn der Graph von f in p einen Gratpunkt hat, und negativ bei einer Bergflanke.



Summiert man die Werte $\gamma(u)$ über alle *u* in einer Nullumgebung auf, so kann man am Vorzeichen ziemlich sicher ablesen, ob in *p* ein Gratpunkt oder ein Flankenpunkt vorliegt. Im Falle eines Flankenpunktes ersetzt man

f(p) durch das Minimum der Werte von f in einer geeigneten Umgebung von p. Dadurch werden die Flanken der Bergrücken steiler. Insgesamt ergibt sich also folgende Operation:

Wähle Umgebungen U_1 und U_2 von Null in \mathbb{Z}^2 und setze

$$\Gamma(p) = \sum_{u \in U_1} (f(p+u) - f(p)) \cdot (f(p-u) - f(p))$$

Ist $\Gamma(p) < 0$, so ersetze f(p) durch min $\{f(p+u) : u \in U_2\}$.

In der Praxis wählt man für U_1 und U_2 kleine, im Nullpunkt zentrierte Quadrate z.B. mit Seitenlängen 5. Die hellen Linien erscheinen dann meist deutlicher und dünner. Ist das Bild verrauscht oder sind die Linien zerfressen, so empfiehlt es sich, vorher eine lineare Gaußglättung durchzuführen. Will man nur kontrastreichere Linien erhalten, so kann man das Kriterium $\Gamma(p) < 0$ für einen Flankenpunkt abändern zu $\Gamma(p) < \theta$, wobei mit der Schwelle $\theta < 0$ in gewissem Sinne die Steilheit der Flanken ausgedrückt wird.

Die Operation kann iteriert werden, um die Linien schärfer hervortreten zu lassen.

Die Hervorhebung von dunklen Linien funktioniert analog. Das Kriterium für Flankenpunkte bleibt natürlich dasselbe; man gibt in den Flankenpunkten lediglich das Maximum von f in U_2 statt des Minimums als Ergebnis aus.

Das folgende Beispiel zeigt links einen Ausschnitt eines Angiographiebildes einer Ratte; man sieht Gefäße, die infolge von Kontrastmittel, das Röntgenstrahlung absorbiert, dunkel erscheinen. Die Bilder rechts daneben zeigen die Ergebnisse nach 1- und 3-maliger Iteration der Hervorhebungsoperation.



Kapitel 4

Detektion von Kanten

4.1 Kontinuierliche Modellierung von Kanten

Das Bildsignal f sei stetig differenzierbar, $\vartheta > 0$, $\|\cdot\|$ eine Norm auf \mathbb{R}^2 .

4.1.1 Gradientenschwellwertentscheidung

Markiere (x, y) als Kantenpunkt, wenn $\|\nabla f(x, y)\| > \vartheta$. Probleme:

- Rauschempfindlichkeit: Auch Störungen mit kleiner Amplitude können große Gradientenlängen bewirken. Deshalb ist es oft unmöglich, eine geeignete Schwelle zu finden, so daß die Ränder kontrastarmer, größerer Strukturen, jedoch nicht die Störungen markiert werden.
- b) Schlechte Lokalisierung: Kanten, deren Grauwertverlauf nicht sprunghaft, sondern rampenförmig verläuft, werden als dicke Kurven markiert, wenn überhaupt. $\square \nabla f \parallel > \vartheta$

Lösungsansätze:

- a) Führe vor der Differentiation eine Glättung durch. Vielfach wurden und werden noch immer lineare Glättungsfilter, meist Gaußfilter, verwendet.
- b) Breite Kantenmarkierungen lassen sich vermeiden, wenn man die Schwellwertentscheidung mit einer Nebenbedingung verbindet:
 - (1) Berechne $\nu(x,y) := \text{Richtung von } \nabla f(x,y) = \frac{\nabla f(x,y)}{\|\nabla f(x,y)\|}.$
 - (2) Setze $\varphi(t) := f((x, y) + t\nu(x, y)).$
 - (3) Markiere (x, y) als Kantenpunkt, wenn $\|\nabla f(x, y)\| > \vartheta$ und wenn $\varphi' = \frac{\partial \varphi}{\partial t}$ in Null ein lokales Maximum hat. (Non maximum suppression) Beachte: Es ist stets $\varphi'(0) \ge 0$, denn $\varphi'(0) = \nu(x, y) \cdot \nabla f(x, y) = \|\nabla f(x, y)\|$.



Bemerkung. Statt lokaler Maxima von φ' kann man natürlich auch Nulldurchgänge von φ'' suchen. Dann hat man einen (nichtlinearen) Differentialoperator 2. Ordnung als Kantenfinder.

4.1.2 Laplace-Nulldurchgänge

Markiere (x, y) als Kantenpunkt, wenn Δf bei (x, y) das Vorzeichen wechselt. Vergleiche 3.2.5. Probleme:

- a) Rauschempfindlichkeit.
- b) Ist $\Delta f : \mathbb{R}^2 \to \mathbb{R}$ stetig differenzierbar mit nicht verschwindendem Gradienten, so ist die folgende Menge $\{(x, y) \mid \Delta f(x, y) = 0\}$ eine eindimensionale Untermannigfaltigkeit und besteht folglich aus Kurven, die kein Ende haben.



Beachte: Kanten in Bildern können einfach irgendwo aufhören.

Lösungsmöglichkeit zu a): Vorschalten eines Glättungsfilters. Verwendet man ein Gaußfilter, so erhält man ein sogenanntes LOG-Filter (Laplace of Gaussian).

4.2 Diskretisierung

Lineare Glättungsfilter, insbesondere Gaußfilter, lassen sich problemlos diskretisieren, indem man die Gewichtsfunktion in den Bildpunkten abtastet und eventuell neu normiert. Der Gradient oder andere Differentialoperatoren (auch zweiter Ordnung) lassen sich nur unvollkommen durch Differenzenoperatoren approximieren.

Die Hintereinanderschaltung eines linearen Gaußfilters und eines linearen Differentialoperators läßt sich hingegen leicht diskretisieren:

Das (isotrope) lineare Gaußfilter im Kontinuierlichen läßt sich schreiben in der Form

$$LG_{\sigma}f(x,y) = g_{2,\sigma} \star f(x,y) := \int_{\mathbb{R}^2} g_{2,\sigma}(u-x,v-y)f(u,v) \, du dv,$$

wobe
i $g_{2,\sigma}(x,y) := \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ die zweidimensionale Gaußfunktion ist. Damit folgt durch Vertauschung von Differentiation und Integration

$$\frac{\partial}{\partial x}LG_{\sigma}f = \frac{\partial g_{2,\sigma}}{\partial x} \star f \quad \text{und} \quad \frac{\partial}{\partial y}LG_{\sigma}f = \frac{\partial g_{2,\sigma}}{\partial y} \star f,$$

und somit

$$\nabla LG_{\sigma}f = \nabla g_{2,\sigma} \star f$$
 sowie $\Delta LG_{\sigma}f = \Delta g_{2,\sigma} \star f.$

Fazit: Die Differentiation läßt sich ganz auf die Gaußfunktion schieben und analytisch exakt berechnen. Zur Diskretisierung ersetzt man das Faltungsintegral durch eine endliche Summe über Abtastwerte. Man muß nur darauf achten, daß die Abtastwerte von $\nabla g_{2,\sigma}$ bzw. $\Delta g_{2,\sigma}$ sich zu Null summieren.

4.3 Cannys Kantenfinder

Kontinuierlich: Sei $g_{2,\sigma}(x,y) := \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$.

- (1) Berechne $\nu(x, y) :=$ Richtung von $\nabla(g_{2,\sigma} \star f)(x, y)$.
- (2) Setze $\varphi_{x,y}(t) := g_{2,\sigma} \star f((x,y) + t\nu(x,y)).$
- (3) Markiere (x, y) als Kantenpunkt, falls $\|\nabla (g_{2,\sigma} \star f)(x, y)\| \ge \vartheta$, und wenn $\varphi'_{x,y}$ in Null ein lokales Maximum hat.

34

4.4. MARR-HILDRETH-KANTENFINDER

Bemerkung. Durch die vorgeschaltete Glättung wird die Detektionssicherheit erhöht, die Lokalisierungspräzision jedoch verschlechtert.

Die Diskretisierung kann analog wie in 4.2 durchgeführt werden.

Problematisch scheint allerdings φ zu sein, weil es in \mathbb{Z}^2 keine Geraden beliebiger Richtung gibt. Canny definiert $\varphi_{x,y}$ lokal bei (x, y), indem er $g_{2,\sigma} \star f$ von den benachbarten Gitterpunkten aus bilinear interpoliert.

Erfahrungen: Cannys Kantenfinder liefert gute Ergebnisse, aber mit wachsendem σ werden die Kanten etwas verschoben und Ecken abgerundet. Manchmal wird eine Kantenverfolgung eingebaut, durch die auch noch Kantenpunkte markiert werden, wenn die Gradientenlänge lokal unter ϑ fällt.

Schwellwertentscheidung mit Hysterese:

Wähle $\vartheta_{\text{unten}} \leq \vartheta_{\text{oben}}$. Sei Z die Menge aller (x, y), für die $\varphi'_{x,y}$ in Null ein lokales Maximum hat. Setze $Z_{\text{oben}} = \{(x, y) \in Z \mid \text{ Gradientenlänge } \geq \vartheta_{\text{oben}}\}$ und $Z_{\text{unten}} = \{(x, y) \in Z \mid \text{ Gradientenlänge } \geq \vartheta_{\text{unten}}\}.$

Als Kantenpunkte werden nun alle diejenigen Pixel in Z_{unten} markiert, die durch einen Weg innerhalb Z_{unten} mit einem Punkt in Z_{oben} verbunden werden können.

Literaturhinweis: [1]

4.4 Marr-Hildreth-Kantenfinder

Ein Vorläufer von Cannys Kantenfinder, der immer noch häufig erwähnt und manchmal auch benutzt wird, ist der von Marr und Hildreth. Er besteht aus einem LOG-Filter, dem ein Nulldurchgangsfinder nachgeschaltet ist. Daß damit Kanten zu finden sind, hatten wir schon in 3.2.5 überlegt. Weil der Ausgang eines LOG-Filters auch dann Null ist, wenn das Eingangssignal keine Kante hat, sondern lokal linear verläuft, genügt es nicht, einfach alle Nullstellen zu markieren; man muß vielmehr die Stellen suchen, in denen das Ausgangsignal des LOG-Filters einen Vorzeichenwechsel hat.

4.5 Bemerkungen

Die Kantenfinder von Canny und Marr-Hildreth sind beide richtungsunabhängig. Während Marr-Hildreth einen *linearen* Differentialoperator, nämlich den Laplace-Operator benutzt, verwendet Canny einen *nichtlinearen* Differentialoperator, durch den ebenfalls die Richtungsunabhängigkeit erreicht wird. Dadurch lassen sich mit Cannys Operator Kanten sehr viel besser finden.

Bei beiden Kantenfindern wird die Entscheidung dadurch robuster gegenüber Rauschen gemacht, daß ein Gaußsches Glättungsfilter vorgeschaltet wird. Andererseits werden dadurch die Kanten auch verschmiert und somit weniger gut lokalisierbar.

Canny hat dazu im eindimensionalen (also für geradlinige Kanten) zwei numerische Kritieren für die Sicherheit der Detektion und für die Präzision ihrer Lokalisierung formuliert. Er zeigte, daß das Produkt der beiden dann am größten ist, wenn das vorgeschaltete Glättungsfilter einem Gaußfilter stark ähnelt. Allerdings führte er diese Optimierung nur über die Menge der linearen Glättungsfilter durch.

4.6 Nichtlineare Differenzenfilter mit Nulldurchgangsdetektoren

Nichtlineare Differenzenfilter $NLH_{\tau,\sigma}$ gemäß 3.2.6 produzieren ähnliche Ergebnisse wie Laplacefilter. Sie umranden die Kanten mit Plus-Minus-Säumen, so daß die Kanten selbst gerade die Nulldurchgänge sind. Sie sind etwas weniger rauschempfindlich als Laplace-Filter.

Der wesentliche Unterschied ist aber, daß die Vorzeichenwechsel im Ausgang längs Kurven vorkommen, die auch an einem Punkt enden können.

4.7 Vorschalten einer kantenerhaltenden Glättung

Unterwirft man ein Bild zunächst einer kantenerhaltenden Glättung, so wird eine nachfolgende Kantendetektion meist recht unproblematisch in folgendem Sinn:

- Welche Kanten gefunden werden, hängt nicht mehr so stark von dem eingesetzten Kantenfinder ab. Selbst einfachste Filter (Sobel u.ä.) liefern oft brauchbare Ergebnisse.
- Die Wahl der Parameter der Kantenfinder ist nicht sehr kritisch, das heißt, die Parameter können in gewissen Grenzen geändert werden, ohne daß sich dadurch die Menge der markierten Kantenpunkte sehr ändert.
- Die Kanten lassen sich sehr präzise lokalisieren.

Einige Verfahren zur kantenerhaltenden Glättung haben bereits einen Kantenfinder mit eingebaut, z.B. das von Blake-Zisserman oder das von Geman-Geman.

Kapitel 5

Binärisierung und Merkmalsextraktion

5.1 Harte und weiche Entscheidungen

Als Binärisierung bezeichnet man den Vorgang, aus einem Bild ein Binärbild zu machen. Im einfachsten Fall wird aus einem Grauwertbild $f: R \to \mathbb{R}$ durch eine Schwellwertentscheidung ein Binärbild $g: R \to \{0, 1\}$ mit g(p) = 1, falls $f(p) \geq \vartheta$, und Null sonst, gemacht. Die Situation kann aber auch viel komplizierter sein:

- Das eingegebene Bild kann ein vektorwertiges Bildsignal $f : R \to \mathbb{R}^n$ sein, dessen Komponenten z.B. Bilder sind, die in verschiedenen Spektralbereichen (Farben) oder mit unterschiedlichen Verfahren aufgenommen wurden.
- Statt einer einzigen Schwellwertentscheidung wird erst eine Reihe von Verarbeitungsschritten durchgeführt, bis schließlich eine Schwellwertentscheidung getroffen wird.

Letztendlich handelt es sich immer um eine Binärisierung, wenn anhand eines gegebenen Bildsignals Objekte oder Bereiche mit gewissen Eigenschaften gefunden und markiert werden sollen. Die Eigenschaften sind meist nicht punktueller Natur, sondern lokaler; das heißt, sie lassen sich nur in genügend großen Bereichen sinnvoll definieren.

Somit genügt es meist nicht, eine simple Schwellwertentscheidung aufgrund des Signalwerts f(p) für jedes Pixel p durchzuführen. Vielmehr versucht man, mit verschiedenen Filteroperationen die jeweils typischen Eigenschaften der lokalen Bildbereiche zu gewinnen.

Meist vermeidet man hier harte Entscheidungen, das heißt, Ja-Nein-Entscheidungen, und bevorzugt weiche Entscheidungen, das heißt, man extrahiert aus dem Bildsignal f an jeder Stelle p einen Zahlenwert, der angibt, wie deutlich das verwendete Filterverfahren eine gewisse Eigenschaft bei p erkennen kann. Man spricht von Merkmalsextraktion.

Solche weiche Entscheidungen sind oft viel robuster gegenüber Störeinflüssen als harte Entscheidungen. Während kleine Änderungen von Störungen des Bildsignals oder des Schwellwertes bewirken können, daß eine harte Entscheidung "umkippt", werden sie bei weichen Entscheidungen meist nur kleine, stetige Änderungen des Ausgangswertes bewirken.

Letztendlich will man aber eine harte Entscheidung treffen. Die vorausgehende weiche Entscheidung oder Merkmalsextraktion sollte so gestaltet sein, daß die nachfolgende harte Entscheidung unempfindlicher gegenüber Störungen der Bilddaten oder Veränderung der Schwellwerte wird. Dies läßt sich manchmal dadurch erreichen, daß man vor die harte Entscheidung eine ganze Kette von weichen Entscheidungen schaltet, die anfangs sehr weich sind und allmählich härter werden und bewirken, daß die der harten Entscheidung angebotenen Eingangsdaten kaum noch in dem sensiblen Bereich um die Schwellwerte herum liegen.

Dieses Vorgehen läßt sich schön am Beispiel der Segmentierung von Bereichen nach Grauwerten demonstrieren.

5.2 Grauwertsegmentierung und Glättung

Sei $f_{\text{ideal}}: R \to \{\alpha_0, \alpha_1\}$ ein Binärbild mit $\alpha_0 < \alpha_1$. Das Bild zerfalle in die beiden Bereiche $B_0 = f_{\text{ideal}}^{-1}(\alpha_0)$ und $B_1 = f_{\text{ideal}}^{-1}(\alpha_1)$. Sei $f = f_{\text{ideal}} + r$ eine additiv gestörte Version.

Ziel: Bestimme aus f die f_{ideal} zugrundeliegende Bereichseinteilung; aber nicht unbedingt die exakten Grauwerte α_0 und α_1 .

Inwieweit dies möglich ist, hängt natürlich stark von r ab.

5.2.1 Einfache Schwellwertentscheidung

Wähle ein $\vartheta \in \mathbb{R}$ und setze $\tilde{B}_0 := \{(x, y) \in R \mid f(x, y) < \vartheta\}$ und $\tilde{B}_1 := \{(x, y) \in R \mid f(x, y) \ge \vartheta\}.$

Wie soll man die Schwelle ϑ wählen, damit $\tilde{B}_0 = B_0$ und $\tilde{B}_1 = B_1$ gilt?

So eine Schwelle ϑ gibt es, falls $|r| < \frac{1}{2}(\alpha_1 - \alpha_0)$ gilt. Dann kann man $\vartheta = \frac{1}{2}(\alpha_0 + \alpha_1)$ setzen. Das Problem ist nur, daß man weder α_0 noch α_1 noch sup |r| kennt.

Schätzung von α_0 und α_1 aus dem Histogramm von f:

Annahme: Das Histogramm sei *bimodal*, das heißt, es habe genau zwei ausgeprägte lokale Maxima, und zwar in $\tilde{\alpha}_0$ und $\tilde{\alpha}_1$, $\tilde{\alpha}_0 < \tilde{\alpha}_1$. Nimmt man ferner an, daß r eine Realisierung eines Rauschprozesses (mit Erwartungswert Null) ist, so sind $\tilde{\alpha}_0$ bzw. $\tilde{\alpha}_1$ mit hoher Wahrscheinlichkeit gute Schätzwerte für α_0 bzw. α_1 .



Als Entscheidungsschwelle ϑ kann man dann $\frac{1}{2}(\tilde{\alpha}_0 + \tilde{\alpha}_1)$ wählen oder vielleicht noch besser eine Stelle zwischen $\tilde{\alpha}_0$ und $\tilde{\alpha}_1$, in der das Histogramm ein tiefes, ausgeprägtes Minimum hat.

Falls |r| größer als $\alpha_1 - \alpha_0$ werden kann, z.B. wenn das Rauschen Gaußsch ist, werden durch solche harten Entscheidungen manche Pixel notwendigerweise falsch klassifiziert, selbst wenn das Histogramm noch bimodal ist.

Ein großer Störterm bewirkt, daß das Histogramm nicht mehr bimodal ist. Dann versagt obige Schwellwertwahl.

Wie man die Schwelle in Abhängigkeit vom Rauschprozeß bestmöglich wählt, wird in der statistischen Entscheidungstheorie untersucht.

5.2.2 Schwellwertentscheidung mit vorgeschalteter linearer Glättung

Durch Anwendung eines linearen Glättungsfilters kann man oft erreichen, daß das Histogramm bimodal wird und eine nachfolgende Schwellwertentscheidung gemäß 5.2.1 unempfindlicher gegenüber Änderungen des Störterms rwird. Dies beruht auf der schon in 3.1.10 untersuchten Verminderung der Rauschvarianz und der Tatsache, daß die Entscheidung nicht nur auf dem einen Wert f(x, y), sondern eben einem lokalen Mittelwert über benachbarte Pixel beruht. Andererseits werden aber gerade dadurch am Rand von B_0 zusätzliche Fehler eingeführt.

5.2.3 Vorschaltung einer nichtlinearen Gaußfilterkette

Ein nichtlineares Gaußfilter kombiniert eine lokale Mittelung mit einer Entscheidung darüber, welche Nachbarpixelwerte in die Mittelung eingehen. Und diese Entscheidung ist weich in dem Sinne, daß ein Nachbarwert f(u, v)von der Mittelung bei (x, y) nicht völlig ausgeschlossen wird, wenn |f(u, v) - f(x, y)| eine vorgegebene Schwelle überschreitet, sondern daß nur seine Gewichtung proportional zu $g_{\eta}(|f(u, v) - f(x, y)|)$ zunimmt. Die Breite η der Gaußglocke spielt die Rolle einer weichen Schwelle.

In einer nichtlinearen Gaußfilterkette nimmt der Parameter η von Filterstufe zu Filterstufe ab. Die Filterkette läßt sich daher als eine Kette von weichen Entscheidungen verstehen, die zunehmend härter wird, weil die Flanken von g_{σ} mit abnehmenden σ steiler werden.

Mit einer nichtlinearen Gaußfilterkette erzielt man oft eine ausgeprägte Bimodularität des Histogramms.

5.3 Lokale Binärisierung kleiner Strukturen

5.3.1 Adaptive Schwellen

Das Bild f_0 enthalte kleinere dunkle Strukturen auf hellerem Hintergrund; man denke z.B. an Text. Häufig besteht der Störterm nicht nur aus Rauschen, sondern hat auch einen langsam variierenden Anteil z.B. infolge ungleichmäßiger Beleuchtung. Dann ist das Histogramm von $f = f_0 + r$ meist nicht bimodal. Eine Segmentierung mit einer globalen Schwellwertentscheidung ist dann nicht möglich. Man kann aber obige Überlegungen bei jeder Stelle des Bildes auf eine lokale Umgebung anwenden, also eine ortsabhängige Schwelle aufgrund lokaler Histogramme wählen.

Oft läßt sich dies vereinfachen, indem man als Schwelle $\vartheta(x, y)$ bei (x, y) einen Mittelwert von f über eine Umgebung von (x, y) wählt. Benutzt man ein Glättungsfilter T_{glatt} , so ergibt sich die Binärisierung

$$B_0 = \{(x,y) \in R \mid f(x,y) < T_{glatt}f(x,y)\} = \{(x,y) \in R \mid (f - T_{glatt}f)(x,y) < 0\} \text{ und}$$
$$\tilde{B}_1 = \{(x,y) \in R \mid (f - T_{glatt}f)(x,y) \ge 0\}.$$

 $f - T_{\text{glatt}} f$ ist das in 3.2.6 betrachtete Hochpaßfilter.

Wenn die zu trennenden helleren und dunkleren Bereiche unterschiedlich groß sind (z.B. Buchstaben auf Hintergrund), ist es günstig, die Schwelle $T_{\text{glatt}}f$ mit einem Faktor $\alpha \approx 1$ zu multiplizieren, also folgende Entscheidung zu treffen: $\tilde{B}_0 = \{(x, y) \in R \mid f(x, y) < \alpha T_{\text{glatt}}f(x, y)\}$ und $\tilde{B}_1 = \{(x, y) \in R \mid f(x, y) \geq \alpha T_{\text{glatt}}f(x, y)\}$.

Häufig wählt man für T_{glatt} ein Gaußsches Glättungsfilter, dessen Standardabweichung σ so gewählt wird, daß 2σ mindestens die Dicke der zu segmentierenden Bildstrukturen ist. Funktionsweise:



Wenn man das lineare Hochpaßfilter Id $-T_{\text{glatt}}$ durch das in 3.2.6 beschriebene nichtlineare $NLH_{\tau,\sigma}$ mit Totbereich $[-\tau, \tau]$ ersetzt, erhält man ein Verfahren zur lokalen Binärisierung, das unempfindlich gegenüber Rauschen mit Amplitude $< \tau$ ist.

5.4 Detektion von Linien

5.4.1 Transversalitätskriterium

Unter Linien verstehen wir dünne Strukturen mit annähernd konstanter Helligkeit, die in einer Richtung deutlich weiter ausgedehnt sind als quer dazu. Im Kontinuierlichen denkt man natürlich an Kurven.

 Γ sei eine Kurve im \mathbb{R}^2 , die glatt ist, d.h. überall eine Tangente besitzt. Ist p ein Punkt auf Γ , so schneidet jede Gerade durch p, deren Richtung nicht tangential ist, die Kurve Γ transversal in p, d.h. p ist ein isolierter Punkt im Durchschnitt $\Gamma \cap G$. Verdickt man die Kurve zu einem Streifen S, so ist die Zusammenhangskomponente des Durchschnitts $S \cap G$, die p enthält, eine Strecke, deren Länge desto kleiner ist, je senkrechter G auf der Tangente von Γ in p steht. Der Mittelpunkt dieser Strecke liegt annähernd in der Mitte des Streifens, und zwar um so genauer, je kürzer die Strecke ist. Um also eine Mittelkurve des Streifens zu markieren, sucht man durch jeden Punkt des Streifens eine Gerade G mit kürzester Schnittstrecke und markiert deren Mittelpunkt.



In realen Bildern ist die Grauwertfunktion innerhalb der Linienstruktur meist nicht wirklich konstant. Wir gehen aber davon aus, dass die lokale Grauwertvariation innerhalb der Linie deutlich kleiner als am Rand ist, so dass man beim Laufen längs einer Geraden das Verlassen der Linien daran erkennen kann, dass die Grauwertänderung eine vorgegebene Schwelle überschreitet.

Detektion von hellen Linien: Wähle Parameter W > 0 und $\theta > 0$. Ein Punkt p liegt auf einer hellen Linie, wenn es eine Richtung r gibt, so dass gilt: Läuft man von p aus auf der Geraden G in Richtung von r, so stößt man innerhalb der Entfernung W auf einen Punkt q mit $f(q) \leq f(p) - \theta$, und ebenso, wenn man in der entgegengesetzten Richtung -r läuft. Sind q_r und q_{-r} die ersten solchen Punkte (also die p am nächsten liegenden), so liegt der Mittelpunkt der Strecke $\overline{q_r q_{-r}}$ annähernd auf der Mitte der Linie.

Mit θ wird also der Kontrast vorgegeben, durch den der Linienrand definiert wird. Und mit W wird eine obere Schranke für die zulässige Lineinbreite angegeben.

Die Übertragung dieser Idee ins Diskrete stößt auf die Schwierigkeit, dass es in einem diskreten Pixelraster keine Geraden beliebiger Richtung gibt. Solange man nur relativ dünne Linienstrukturen betrachtet, genügt es aber meist, lediglich achsenparallele und diagonale Geraden G zu verwenden.

Die Detektion von dunklen Linien oder von Linien, deren Umgebung auf der einen Seite heller und auf der anderen dunkler ist, lässt sich analog formulieren.







5.4. DETEKTION VON LINIEN

In dem obigen Röntgenbild wurden die dicken weißen Adern nicht markiert, weil der Parameter W für die Weite zu klein war. Das Bild wurde erst in der Größe verdoppelt und mit einem nichtlinearen Gaußfilter mit $\sigma = 2$ und $\eta = 16$ geglättet; dann wurden mit W = 4 und $\theta = 2$ die Linien markiert.

Achtung! Der oben skizzierte Algorithmus überprüft nur, ob die Struktur in mindestens einer Richtung dünn ist, jedoch nicht, ob sie in der Querrichtung weiter ausgedehnt ist, wie man es von einer Linie eigentlich erwartet. Er markiert also auch kleine Flecken mit lokal maximaler Helligkeit, wie man auch an obigen Beispielen sieht. Bei welchem Ausdehnungverhältnis eine Trennung zwischen Linie und Flecken gezogen wird, muss der Anwender durch eine Nachbearbeitung festlegen. Man kann z.B. die Anzahl der Pixel in einer markierten Struktur als Kriterium nehmen. Manchmal hilft auch, vor der Linienmarkierung eine (eventuell kantenerhaltende) Glättung oder eine Verdünnung der Linien durchzuführen.

Die folgenden Bilder zeigen links das Originalbild und in der Mitte das Ergebnis der Liniendetektion, der eine lineare Gaußfilterung mit $\sigma = 2$ vorgeschaltet wurde. Das Bild rechts entstand dadurch, dass kleine schwarze Linienstücke mit weniger als 50 Pixeln weggelassen wurden.



5.4.2 Gradientenflusskriterium

Im Kontinuierlichen zeigt der Gradient ∇f der Grauwertfunktion f an jeder Stelle in die Richtung des stärksten Anstiegs. Hat der Graph von f über p einen Gipfel- oder einen Gratpunkt, so gibt es eine in p zentrierte Kreislinie K und ein $a \in K$, so dass die Gradienten von f in den Punkten $q_1 = p + a$ und $q_2 = p - a$ annähernd auf pzuweisen. Um dies zu präzisieren, setzen wir $\varphi(q_1) = -||a||^{-1} < a|\nabla f(q_1) > und \varphi(q_2) = ||a||^{-1} < a|\nabla f(q_2) >$. Diese beiden Größen kann man als lokalen Fluss in Richtung von p deuten. Für geeignetes a sind beide deutlich positiv, wenn p in einer dünnen Struktur liegt.



Im Diskreten hat man keine richtigen Kreislinien zur Verfügung. Deshalb lässt man a über die Seitenlinien eines Quadrates mit Seitenlänge 2W + 1 laufen. Zuvor muss man noch in jedem Pixel p eine Gradientschätzung berechnen; dafür eignet sich z.B. die baryzentrische Gradientenschätzung in einem 3×3 -Quadrat (siehe 2.6). Ein Pixel p wird als ein Gratpunkt markiert, wenn es ein a gibt, so dass $\varphi(q_1)$ und $\varphi(q_2)$ beide positiv sind und ihre Summe eine vorgegebene Schranke ϕ übersteigt.

in dem folgenden Bild wurden mit der Gradientflussmethode mit W = 2 und $\phi = 60$ die Linien markiert und dann Zusammenhangskomponenten mit weniger als 10 Pixeln weggelassen.



Das Bild unten links wurde mit einem linearen Gaußfilter mit $\sigma = 2$ geglättet; anschließend wurden mit der Gradientenflussmethode (mit W = 1 und $\phi = 10$ die Adern markiert und dann kleinere Zusammenhangskomponenten weggelassen.



In dem folgenden Röntgenbild wurden ohne weitere Vorverarbeitung mit der Gradientenflussmethode (mit W = 1 und $\phi = 60$) die Adern markiert.

5.4. DETEKTION VON LINIEN



Die Markierung der Adern fällt leichter, wenn das Bild zunächst mit bilinearer Interpolation auf die doppelte Größe gebracht wird. Anschließend wurde die Gradientenflussmethode mit W = 2 und $\phi = 30$ angewendet, um die Adern zu markieren, und dann wurden kleine Adernstücke mit weniger als 20 Pixeln weggelassen.



5.4.3 Bemerkung

Für den menschlichen Betrachter sehen Bildstrukturen oft gefälliger und deutlicher aus, wenn sie noch nicht richtig binärisiert sind, weil dann ihre Ränder glatter erscheinen. Man sollte sich dadurch aber nicht täuschen lassen und eventuell die Leistungsfähigkeit von Algorithmen falsch beurteilen! Beispielsweise sehen in dem Bild links unten die Adern schöner aus als in dem rechten; die beiden Bilder unterscheiden sich jedoch nur dadurch, dass im linken Bild die Schwarzweiß-Übergänge durch Einfügen grauer Pixel weicher gestaltet wurde.





5.4.4 Bestimmung der Linienrichtung

Hat man wie in 5.4.1 eine zur Linienstruktur L im Punkt p transversale Gerade G mit möglichst kleinem Durchschnitt gefunden, so kann man recht einfach die Richtung der Linie an dieser Stelle schätzen. Allerdings liefert die naheliegende Idee, einfach eine zu G orthogonale Richtung zu wählen, meist zu ungenaue Ergebnisse, insbesondere wenn man für G im Diskreten nur Parallelen zu den Achsen und den Hauptdiagonalen zulässt. Robuster ist folgendes Vorgehen:

Wähle ein kleines Quadrat U mit Zentrum p. Die Gerade G zerteilt U in zwei Teile U_1 und U_2 . Berechne den Schwerpunkt s_1 von $U_1 \cap L$ und den Schwerpunkt s_2 von $U_2 \cap L$. Die Verbindungsstrecke $\overline{s_1 s_2}$ verläuft ziemlich parallel zu L, sofern L bei p nicht stark gekrümmt ist.



Man kann das Verfahren in 5.4.1 dadurch abändern, dass man nicht p markiert, sondern gleich die ganze Strecke $\overline{s_1 s_2}$ einzeichnet. Das Bild unten rechts ist auf diese Weise entstanden, wobei eine lineare Gaußglättung mit $\sigma = 2$ vorgeschaltet und eine Eliminierung kleiner Zusammenhangskomponenten mit weniger als 80 Pixeln nachgeschaltet wurde. Als Umgebung U wurde ein 9 × 9-Quadrat gewählt.





Kapitel 6

Verarbeitung von Binärbildern

Ein Binärbild f mit den beiden Grauwerten α_0 und α_1 ist bestimmt durch $K := f^{-1}(\alpha_0)$ und den Definitionsbereich R von f. Für theoretische Betrachtungen wählen wir $R = \mathbb{R}^2$ oder $R = \mathbb{Z}^2$, und wir setzen voraus, daß $K \subset \mathbb{R}^2$ kompakt bzw. $K \subset \mathbb{Z}^2$ beschränkt ist. Wir werden K selbst einfach als *Binärbild* oder *Binärobjekt* bezeichnen, weil es in Anwendungen oft die Silhouette von Objekten ist.

6.1 Topologische Eigenschaften von Binärbildern

Wir betrachten zunächst den kontinuierlichen Fall $R = \mathbb{R}^2$. Sei $K \subset \mathbb{R}^2$ kompakt. Eine Eigenschaft heißt topologisch, wenn sie unter Homöomorphismen invariant ist. **Beispiele.**

z(K) = Anzahl der Zusammenhangskomponenten von K.

 $c(K) = Anzahl der beschränkten Zusammenhangskomponenten von <math>\mathbb{R}^2 \setminus K$. (Löcher von K)

e(K) = z(K) - c(K) = Eulerzahl von K.

Diese Größen reichen manchmal schon aus, um zumindest prinzipiell im kontinuierlichen Bild die in einer Anwendung vorkommenden Binärobjekte zu unterscheiden. Bei der Übertragung ins Diskrete treten jedoch Probleme auf.

6.1.1 Probleme mit dem Zusammenhangsbegriff bei diskreten Binärobjekten

Als Unterraum von \mathbb{R}^2 (mit der euklidischen Topologie) trägt \mathbb{Z}^2 die diskrete Topologie; somit ist jede Teilmenge von \mathbb{Z}^2 offen und abgeschlossen, und jede Teilmenge mit mindestens zwei Elementen ist unzusammenhängend.

Deshalb verwendet man statt des topologischen Umgebungsbegriffs meist einen Ad-hoc-Begriff von Nachbarschaft, der es erlaubt, den Begriff des Wegzusammenhangs im \mathbb{R}^2 in gewissem Umfang nach \mathbb{Z}^2 zu übertragen.

4-Nachbarschaft:

$$\frac{|\times|}{|\times||x,y|\times|} \qquad \qquad N_4(x,y) := \{(u,v) \in \mathbb{Z}^2 \mid |u-x|+|v-y| \le 1\} \\ = \text{Kreisscheibe um } (x,y) \text{ mit Radius 1 bezüglich der } l_1\text{-Norm.}$$

8-Nachbarschaft:

$$\frac{|X| \times |X|}{|X| \times |X|} \qquad \qquad N_8(x, y) := \{(u, v) \in \mathbb{Z}^2 \mid \max\{|u - x|, |v - y|\} \le 1\} \\ = \text{Kreisscheibe um } (x, y) \text{ mit Radius 1 bezüglich der } l_{\infty}\text{-Norm.}$$

Es ist nun naheliegend, eine Folge $p_0, \ldots, p_n \in \mathbb{Z}^2$ als Weg zu bezeichnen, wenn stets $p_{j+1} \in N_4(p_j)$ (bzw. $p_{j+1} \in N_8(p_j)$) gilt. Mit dieser Definition tritt jedoch folgendes Problem auf:



- 1) Verwendet man N_8 , so bilden die mit \times gekennzeichneten Pixel eine geschlossene Kurve, deren Komplement zusammenhängt. Das heißt: Der Jordan'sche Kurvensatz gilt nicht.
- 2) Verwendet man N_4 , so bilden die mit \times gekennzeichneten Pixel eine total unzusammenhängende Menge, deren Komplement zwei Zusammenhangskomponenten hat.

Das ist auch etwas "seltsam".

6.1.2 Definition von Zusammenhang

Man kann obige Probleme vermeiden, indem man den Begriff der Nachbarschaft für Pixel $p \in K$ und für Pixel $p \notin K$ unterschiedlich definiert.

6.1.2.1 Definition. Zu einem gegebenen Binärobjekt $K \subset \mathbb{Z}^2$ seien die Nachbarschaften N(p) der Pixel $p \in \mathbb{Z}^2$ folgendermaßen definiert:

$$N(p) = \begin{cases} N_8(p) & \text{für } p \in K \\ N_4(p) & \text{für } p \notin K \end{cases}$$

Bemerkung. Man kann die beiden Fälle auch vertauschen. Meist ist es aber angemessener, wenn diagonal benachbarte Pixel in K auch als Nachbarn gelten.

6.1.2.2 Definition. Sei ein Binärobjekt $K \subset \mathbb{Z}^2$ gegeben.

- a) Ein Weg in K (bzw. $\mathbb{Z}^2 \setminus K$) ist eine Folge von Pixeln p_0, \ldots, p_n in K (bzw. $\mathbb{Z}^2 \setminus K$), so daß $p_{j+1} \in N(p_j)$. Er verbindet also p_0 mit p_n .
- b) Für $p, q \in \mathbb{Z}^2$ bedeute $p \sim q$, daß entweder p und q durch einen Weg in K oder durch einen Weg in $\mathbb{Z}^2 \setminus K$ verbunden werden können.

6.1.2.3 Lemma. ~ ist eine Äquivalenzrelation auf \mathbb{Z}^2 . Sowohl K als auch $\mathbb{Z}^2 \setminus K$ sind Vereinigungen von Äquivalenzklassen.

Diese Äquivalenzklassen nennen wir die Zusammenhangskomponenten von K bzw. $\mathbb{Z}^2 \setminus K$.

6.1.2.4 Bemerkung. Man kann K (und auch $\mathbb{Z}^2 \setminus K$) als ungerichteten Graph mit Knotenmenge K (bzw. $\mathbb{Z}^2 \setminus K$) und Kantenmenge $E = \{(p,q) \mid q \in N(p)\}$ auffassen. Die Zusammenhangskomponenten dieses Graphen sind die gleichen wie die oben definierten.

6.1.2.5 Lemma. Sei $K = \{p_0, \ldots, p_{n-1}\} \subset \mathbb{Z}^2$. Die Folge (p_0, \ldots, p_{n-1}) sei ein einfach geschlossener Weg, d.h. die p_j seien paarweise verschiedenen und $K \cap N(p_j) = \{p_{j-1}, p_j, p_{j+1}\}$ für jedes j > 0, wobei $p_n := p_0$. Dann hat $\mathbb{Z}^2 \setminus K$ genau zwei Zusammenhangskomponenten , von denen genau eine beschränkt ist.

Bemerke: In diesem Sinn gilt also der Jordansche Kurvensatz. Damit definiert man z(K), c(K) und e(K) wie im kontinuierlichen Fall.

Beweis. Fasse $p_0, \ldots, p_{n-1}, p_0$ als Eckenfolge eines einfach geschlossenen Polygonzugs im \mathbb{R}^2 auf und wende den Jordanschen Kurvensatz an.

6.1.3 Etikettierung (Labeling) von Zusammenhangskomponenten

Aufgabe: Färbe die Zusammenhangskomponenten eines Binärobjektes $K \subset \mathbb{Z}^2$ mit unterschiedlichen Farben ein, oder anders gesagt:

Finde eine Abbildung $\varphi: K \to \mathbb{N}$, so daß für jedes $n \in \mathbb{N}$ die Faser $\varphi^{-1}(n)$ leer oder eine Zusammenhangskomponente von K ist.

Lösungansätze: Nach 6.1.2.4 trägt K eine Graphenstruktur mit den gleichen Zusammenhangskomponenten. Deshalb kann man die üblichen Graphenalgorithmen anwenden, die mit Tiefensuche oder Breitensuche arbeiten. Aufgrund der besonderen Situation kann man auch schnellere Ad-hoc-Algorithmen entwerfen.

6.2 Geometrische Eigenschaften zusammenhängender Binärobjekte

6.2.1 Der kontinuierliche Fall

 $K \subset \mathbb{R}^2$ sei kompakt und zusammenhängend. Eine Eigenschaft von K heißt geometrisch, wenn sie unter den Bewegungen des \mathbb{R}^2 invariant ist.

(Zur Erinnerung: Eine *Bewegung* ist eine Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$ der Form $p \mapsto Ap + b$, wobei $A \in SO(2)$ eine Drehung und $b \in \mathbb{R}^2$ sind.) Beispiele.

- a) Fläche, maximaler Durchmesser, Umfang (sofern definiert, also z.B. wenn K stückweise glatt berandet). Denn Bewegungen sind Isometrien.
- b) Konvexität.

6.2.2 Der diskrete Fall

Beim Übergang vom Kontinuierlichen zum Diskreten treten folgende Probleme auf:

1) Die Bewegungen des \mathbb{R}^2 induzieren keine Operation auf \mathbb{Z}^2 (mit wenigen Ausnahmen, wie Rotationen um Vielfache von 90°, Translationen um ganzzahlige Vektoren).

Insbesondere bedeutet dies: Bewegungen kommutieren nicht mit der Diskretisierung.

Oder anders ausgedrückt: Bewegt man ein Objekt $K \subset \mathbb{R}^2$, so wird sein diskretisiertes Binärbild einer Transformation unterzogen, die meist von der Bewegung verschieden ist.

2) Begriffe wie Fläche, Umfang und Konvexität sind im Diskreten zunächst nicht definiert. Häufig diskretisiert man die entsprechenden Formeln im \mathbb{R}^2 auf irgendeine Weise; z.B. nimmt man die Anzahl der Pixel in K als Ersatz für die Fläche. Problematischer ist es schon, einen Ersatz für den Umfang oder den Begriff der Randkurve zu finden.

Wie auch immer man die Definitionen trifft, man wird wegen 1) nicht erreichen, daß die so definierten Größen invariant unter Bewegungen des realen Objekts vor der Kamera sind. Man kann höchstens erreichen, daß sie sich nur wenig ändern, indem man die Pixelauflösung des Bildes groß wählt im Vergleich zu den Details der Gestalt des Objekts.

Bemerkung. Es gibt natürlich Ansätze, von vornherein Geometrie und Topologie in dem diskreten Gitter \mathbb{Z}^2 zu entwickeln. Nur beschreibt man damit nicht unbedingt die Situation in Anwendungen, in denen die Binärobjekte die diskretisierten Silhouetten von realen Objekten vor der Kamera sind.

Man überlege sich einmal, wie man den Umfang diskreter Binärobjekte definieren kann, so daß der wirkliche Umfang eines mit der Kamera aufgenommenen Objekts gut approximiert wird und keine allzu großen Änderungen auftreten, wenn es im \mathbb{R}^2 bewegt wird.

6.3 Bestimmung der Lage und Position von Binärobjekten

Aufgabe: Gegeben seien eine kompakte Menge $K \subset \mathbb{R}^2$ mit positivem Lebesgue-Maß und das Bild K_1 von K unter einer Bewegung φ des \mathbb{R}^2 . Bestimme φ , d.h. bestimme eine Drehung $A \in SO(2)$ und ein $b \in \mathbb{R}^2$, so daß $K_1 = A(K) + b$.

Dafür gibt es mehrere, ganz unterschiedliche Lösungsansätze. Wir beschreiben zunächst die sogenannte Momentenmethode, die relativ einfach ist, aber trotzdem oft ausreichend gute Ergebnisse liefert.

Alle Methoden beruhen darauf, daß kompakten Mengen im \mathbb{R}^2 (eventuell mit zusätzlichen Eigenschaften) Größen zugeordnet werden, aus deren Änderungen unter einer Bewegung man Rückschlüsse auf die Bewegung ziehen kann. Für praktische Anwendungen wird dann die Definition dieser Größen in geeigneter Weise diskretisiert.

6.3.1 Die Momentenmethode

0-tes Moment:

$$M_K := \iint_K 1 \, dx dy.$$

 M_K wird *Fläche* (oder *Masse*) von *K* genannt. 1-te Momente:

$$\overline{x}_K := \frac{1}{M_K} \iint_K x \, dx dy \qquad \text{und} \qquad \overline{y}_K := \frac{1}{M_K} \iint_K y \, dx dy$$

 $(\overline{x}_K, \overline{y}_K)$ wird *Schwerpunkt* von *K* genannt. 2-te Momente:

$$\begin{split} \Theta_{xx}^{K} &:= \iint_{K} (x - \overline{x}_{K})^{2} \, dx dy \\ \Theta_{xy}^{K} &:= \Theta_{yx}^{K} := \iint_{K} (x - \overline{x}_{K})(y - \overline{y}_{K}) \, dx dy \\ \Theta_{yy}^{K} &:= \iint_{K} (y - \overline{y}_{K})^{2} \, dx dy \\ \Theta_{K} &:= \left(\begin{array}{c} \Theta_{xx}^{K} & \Theta_{xy}^{K} \\ \Theta_{yx}^{K} & \Theta_{yy}^{K} \end{array} \right) = \iint_{K} \left(\begin{array}{c} x - \overline{x}_{K} \\ y - \overline{y}_{K} \end{array} \right) (x - \overline{x}_{K}, y - \overline{y}_{K}) \, dx dy \end{split}$$

 Θ_K wird Trägheitsmatrix genannt. Die Trägheitsform T_K sei die quadratische Form

$$T_{K}(u,v) := \Theta_{xx}^{K} u^{2} + 2\Theta_{xy}^{K} uv + \Theta_{yy}^{K} v^{2} = (u,v) \Theta_{K}(u,v)^{T}.$$

Anschauliche Interpretation:

Fasse K als homogene Metallplatte (der Dicke 1) auf. Dann ist M_K die Masse von K und $(\overline{x}_K, \overline{y}_K)$ tatsächlich der Schwerpunkt.

 $T_K(u, v)$ ist (bis auf den Faktor $\frac{1}{2}$) die Rotationsenergie, wenn die Metallplatte K mit Winkelgeschwindigkeit ||(u, v)|| um die Gerade $G = (\overline{x}_K, \overline{y}_K) + \mathbb{R}(u, v)^{\perp}$ rotiert.

Das sieht man folgendermaßen:

$$T_{K}(u,v) = \iint_{K} \left((x - \overline{x}_{K})^{2} u^{2} + 2(x - \overline{x}_{K})(y - \overline{y}_{K})uv + (y - \overline{y}_{K})^{2} v^{2} \right) dxdy$$

$$= \iint_{K} \left((x - \overline{x}_{K})u + (y - \overline{y}_{K})v \right)^{2} dxdy$$

$$= \iint_{K} \left\langle (u,v), (x - \overline{x}_{K}, y - \overline{y}_{K}) \right\rangle^{2} dxdy$$

$$= ||(u,v)||^{2} \cdot \iint_{K} (\text{Abstand von } (x,y) \text{ zu } G)^{2} dxdy$$
Trägheitsmoment von K

6.3.1.1 Lemma. $K \subset \mathbb{R}^2$ sei kompakt; φ sei eine Bewegung des \mathbb{R}^2 der Gestalt $\varphi(x, y) = A(x, y)^T + b$ mit $A \in SO(2)$ und $b \in \mathbb{R}^2$. Dann gilt:

- a) $\varphi(\overline{x}_K, \overline{y}_K)$ ist der Schwerpunkt von $\varphi(K)$.
- b) $\Theta_{\varphi(K)} = A \cdot \Theta_K \cdot A^{-1}$. Insbesondere ist Θ_K invariant unter Translationen.

Beweis. φ ist eine affine, lineare Abbildung, deren Ableitung überall gleich A ist; die Jacobi-Determinante ist also konstant gleich 1. Mit dem Integraltransformationssatz folgt

$$\begin{split} M_{\varphi(K)} &= \iint_{\varphi(K)} 1 \, d\xi d\eta = \iint_{K} |\det(A)| \, dx dy = M_{K} \quad \text{und} \\ (\overline{x}_{\varphi(K)}, \overline{y}_{\varphi(K)}) &= \frac{1}{M_{\varphi(K)}} \iint_{\varphi(K)} (\xi, \eta) \, d\xi d\eta \quad = \quad \frac{1}{M_{K}} \iint_{K} \varphi(x, y) \, dx dy = \varphi \left(\frac{1}{M_{K}} \iint_{K} (x, y) \, dx dy \right) = \varphi(\overline{x}_{K}, \overline{y}_{K}) \\ \Theta_{\varphi(K)} &= \quad \iint_{\varphi(K)} \left(\begin{array}{c} \xi - \varphi_{1}(\overline{x}_{K}, \overline{y}_{K}) \\ \eta - \varphi_{2}(\overline{x}_{K}, \overline{y}_{K}) \end{array} \right) (\xi - \varphi_{1}(\overline{x}_{K}, \overline{y}_{K}), \eta - \varphi_{2}(\overline{x}_{K}, \overline{y}_{K})) \, d\xi d\eta \\ &= \quad \iint_{K} (\varphi(x, y) - \varphi(\overline{x}_{K}, \overline{y}_{K}))(\varphi(x, y) - \varphi(\overline{x}_{K}, \overline{y}_{K}))^{T} |\det(A)| \, dx dy \\ &= \quad \iint_{K} A \left(\begin{array}{c} x - \overline{x}_{K} \\ y - \overline{y}_{K} \end{array} \right) (x - \overline{x}_{K}, y - \overline{y}_{K}) A^{T} \, dx dy \\ &= \quad A \cdot \Theta_{K} \cdot A^{T} \\ &= \quad A \cdot \Theta_{K} \cdot A^{-1} \end{split}$$

6.3.1.2 Korollar. Θ_K und $\Theta_{\varphi(K)}$ haben dieselben beiden Eigenwerte $\lambda_1, \lambda_2 \in \mathbb{R}$. Ist V_j der Eigenraum von Θ_K zu λ_j , so ist $A(V_j)$ der Eigenraum von $\Theta_{\varphi(K)}$ zu λ_j .

Beweis. Θ_K ist eine symmetrische, reelle Matrix und hat deshalb reelle Eigenwerte. Der Rest folgt, weil $\Theta_{\varphi(K)}$ zu Θ_K ähnlich ist.

6.3.1.3 Lage- und Positionsbestimmung

Gegeben seien K und $\varphi(K)$. Zu bestimmen ist die Bewegung φ . Sie habe die Gestalt $\varphi(x, y) = A(x, y)^T + b$ mit $A \in SO(2), b \in \mathbb{R}^2$. Es gilt:

$$\begin{split} \varphi(x,y) &= A \left(\begin{array}{c} x \\ y \end{array} \right) + b &= A \left(\begin{array}{c} x - \overline{x}_K \\ y - \overline{y}_K \end{array} \right) + A \left(\begin{array}{c} \overline{x}_K \\ \overline{y}_K \end{array} \right) + b \\ &= A \left(\begin{array}{c} x - \overline{x}_K \\ y - \overline{y}_K \end{array} \right) + \left(\begin{array}{c} \overline{x}_{\varphi(K)} \\ \overline{y}_{\varphi(K)} \end{array} \right) \\ &= \underbrace{A \left(\begin{array}{c} x - \overline{x}_K \\ y - \overline{y}_K \end{array} \right) + \left(\begin{array}{c} \overline{x}_K \\ \overline{y}_K \end{array} \right)}_{\text{Rotation } \tilde{A} \text{ um den Schwerpunkt von } K} + \left(\begin{array}{c} \overline{x}_{\varphi(K)} \\ \overline{y}_{\varphi(K)} \end{array} \right) - \left(\begin{array}{c} \overline{x}_K \\ \overline{y}_K \end{array} \right), \quad \text{wegen 6.3.1.1.a)} \end{split}$$

Wenn man die Schwerpunkte von K und $\varphi(K)$ berechnet hat, braucht man also nur noch den Drehwinkel von \tilde{A} zu bestimmen.

Voraussetzung: Die Eigenwerte von Θ_K sind verschieden.

Dann sind die Eigenräume eindimensional; sei o.E. V_1 der Eigenraum zum größeren Eigenwert λ_1 . Nach 6.3.1.2 ist dann $A(V_1)$ der Eigenraum von $\Theta_{\varphi(K)}$ zum Eigenwert λ_1 .

Man braucht also nur die Eigenräume von Θ_K und $\Theta_{\varphi(K)}$ zum Eigenwert λ_1 zu berechnen; das sind zwei Geraden, und der Winkel zwischen ihnen ist gleich dem Drehwinkel von A, allerdings nur bis auf $\pm 180^{\circ}$.

Diese Mehrdeutigkeit kann man meist durch Ad-hoc-Methoden eliminieren, z.B. wenn die auf den Eigenräumen liegenden Randpunkte von K nicht symmetrisch verteilt sind.



Die Berechnung der Eigenwerte und Eigenvektoren ist unproblematisch, weil nur quadratische und lineare Gleichungen vorkommen.

6.3.1.4 Bemerkung.

- 1) Sind λ_1 und λ_2 verschieden, so sind die Eigenräume V_1 und V_2 orthogonal zueinander.
- 2) Sind λ_1 und λ_2 gleich, so ist ganz \mathbb{R}^2 der Eigenraum, und obige Methode ist nicht anwendbar. Das ist in praktischen Anwendungen aber selten.
- 3) Die Übertragung der Momentenmethode ins Diskrete ist relativ unproblematisch, wenn die Pixelauflösung hoch ist. Man ersetzt einfach die Integrale durch Summen über die Pixel von K.

6.4 Konturverfolgung

 $K \subset \mathbb{Z}^2$ sei ein zusammenhängendes Binärobjekt ohne Loch mit $\#K \ge 2$. $\partial K = \{p \in K \mid N(p) \cap (\mathbb{Z}^2 \setminus K) \neq \emptyset\}$ bezeichne den Rand von K.

Aufgabe: Stelle ∂K als Weg dar, d.h. finde einen geschlossenen Weg p_0, \ldots, p_n , so daß $\partial K = \{p_0, \ldots, p_n\}$. Versuche *n* möglichst klein zu wählen, also möglichst wenig Pixel doppelt zu durchlaufen.

6.4.1 Erster Lösungsansatz

$$\begin{split} & \text{Wähle ein } p_0 \in \partial K \text{ und ein } q_0 \in N(p_0) \cap (\mathbb{Z}^2 \backslash K); \\ & j = 0; \\ & \text{do } \{ \\ & \text{,numeriere die Nachbarn von } p_j, \text{ beginnend mit } q_0, \text{ entgegen dem Uhrzeigersinn als } q_0, \dots, q_7"; \\ & m = \min\{i \mid q_i \in K\}; \\ & j = j + 1; \\ & p_j = q_m; \\ & q_0 = q_{m-1}; \\ \} \\ & \text{while } p_j \neq p_0; \end{split}$$

Algorithmus 6.1: "überprüft" die direkten Nachbarn

Bemerkungen.

1) Obiger Algorithmus liefert einer Pixelfolge p_0, \ldots, p_n mit den folgenden Eigenschaften:

 $p_{j+1} \in N_8(p_j), \quad p_j \in \partial K \quad \text{und} \quad p_{j+1} \neq p_j.$

2) Die konstruierte Pixelfolge enthält nicht immer alle Randpixel von K:



Das bedeutet, der Algorithmus terminiert nicht immer.

3) Die konstruierte Pixelfolge p_0, \ldots, p_n ist nicht immer injektiv:

	p_4		
p_0	<i>P</i> ₁ =	p_3	
q_0		p_2	

Das passiert nicht, wenn K überall mehr als 1 Pixel dick ist.

6.5. SKELETTIERUNG

6.4.2 Zweiter Lösungsansatz (Turtle Algorithmus)

Anschauliche Idee: Spur einer Schildkröte, die ihre bisherige Richtung um 90° nach links ändert, wenn sie auf ein Pixel außerhalb von K stößt, und um 90° nach rechts, sonst.

Dies ergibt eine mäanderförmige Kurve entlang der Kontur:



Wähle ein $p_0 \in K$, das einen 4-Nachbarn $q \notin K$ in einer Richtung $m \in \{0, 1, 2, 3\}$ hat; j = 1; do { if $(q \in K)$ { m = m - 1 modulo 4; $p_j = q$; j = j + 1; } else m = m + 1 modulo 4; q = 4-Nachbar von q in Richtung m; } while $p_j \neq p_0$; Richtungsnumerierung: $2 \xrightarrow{l} 0$ 3



Bemerkungen.

1) Der Turtle-Algorithmus liefert einer Pixelfolge p_0, \ldots, p_n mit den folgenden Eigenschaften:

$$p_{i+1} \in N_8(p_i)$$
 und $p_i \in \partial K$.

2) Die konstruierte Pixelfolge enthält nicht immer alle Randpixel von K.



3) Die konstruierte Pixelfolge p_0, \ldots, p_n ist nicht immer injektiv; es kann sogar $p_1 = p_0$ vorkommen, was zu einem vorzeitigen Terminieren führt. Dies kann man vermeiden, indem man als **Startpunkt** p_0 die oberste linke Ecke von K wählt, in Bildschirmkoordinaten also $p_0 = (x_0, y_0)$ mit $y_0 = \min\{y: \text{ es gibt } (x, y) \in K\}$ und $x_0 = \min\{x: \text{ es gibt } (x, y_0) \in K\}$, sowie als **Anfangsrichtung** abwärts, also m = 3.

6.5 Skelettierung

K sei ein Binärobjekt.

Aufgabe: "Verdünne" K zu endlich vielen Wegen $\Gamma_1, \ldots, \Gamma_n$, ohne die Anzahl der Zusammenhangskomponenten oder der Löcher zu ändern.

Typisches Beispiel: K ist ein linienhaftes Objekt und die Breite der Linien ist größer als ein Pixel und soll auf ein Pixel reduziert werden. Verjüngung binärisierter Buchstaben oder von Kanten.

Ansatz im Kontinuierlichen, also für $K \subset \mathbb{R}^2$:

Skelett(K) := {
$$p \in K \mid \exists q_1, q_2 \in \partial K, q_1 \neq q_2 : d(p, q_1) = d(p, q_2) = d(p, \partial K)$$
}.

Ist d die euklidische Metrik und ist K stückweise glatt berandet, so ist das Skelett von K die Vereinigung von endlich vielen Kurvenstücken. Beispiele.

Diese Definition des Skeletts läßt sich zwar ins Diskrete übertragen, liefert aber nicht die gewünschten Ergebnisse, weil man keinen isotropen Abstandsbegriff, d.h. rotationsinvariante Kreisscheiben hat. Deshalb versucht man eine Skelettierung durch Verdünnungsalgorithmen zu erreichen, die iterativ geeignete Randpixel entfernen, ohne die Anzahl der Löcher zu verändern oder das Binärobjekt allzu sehr zu schrumpfen. Es gibt mehrere adhoc konstruierte Algorithmen. Wir präsentieren einen, in dem nur lokale Bedingungen in einer Nachbarschaft überprüft werden.

6.5.1 Bedingungen

p darf nicht entfernt werden in folgenden Situationen:

1) p ist Endpunkt eines Weges:



2) p verbindet mehrere Zusammenhangskomponenten seiner Nachbarn, insbesondere p liegt auf einem Weg:



3) p hat nur einen Nachbarn im Komplement:



Denn es sollen keine Kerben im Skelett entstehen und überflüssige kurze oder gekrümmte Skelettstücke vermieden werden.

6.5.2 Verdünnungsalgorithmus

Sei $K \subset \mathbb{Z}^2$ ein Binärobjekt. Die 8-Nachbarn eines Pixels $p \in \mathbb{Z}^2$ seien folgendermaßen durchnumeriert:

p_3	p_2	p_1
p_4	р	p_0
p_5	p_6	p_7

Der folgende Skelettierungsalgorithmus nimmt als Eingabe ein Binärobjekt $K \subset \mathbb{Z}^2$ und gibt eine Skelett Sfür K aus. Innerhalb des Algorithmus werden alle Teilmengen von \mathbb{Z}^2 durch ihre charakteristischen Funktionen beschrieben; insbesondere gilt also $\chi_K(p) = 1$, wenn $p \in K$, und 0 sonst. Der Algorithmus besteht aus einer Schleife, in der K sukzessive bis zu einem Skelett verdünnt wird. Die Schleife stoppt erst, wenn sich keine Veränderungen mehr ergeben. Der Schleifenkörper enthält zwei Verdünnungsschritte, die hintereinander ausgeführt werden und das momentane Binärobjekt von unterschiedlichen Richtungen aus abnagen. g_1 ist das Ergebnis des ersten Schrittes, g_2 das des zweiten Schrittes.

52

```
Eingabe: Eine beschränkte Menge K \subset \mathbb{Z}^2
Ausgabe: Ein Skelett S für K
g_2 = \chi_{\kappa};
do {
   g_0 = g_2
   for (p \in \mathbb{Z}^2) {
                                                                                                             /* 1. Schritt */
      if (g_0(p) = 1
          \& 2 \le \sum_{i=0}^{7} g_0(p_i) \le 6
                                                                                                                           (a)
          & 1 = Anzahl der 0-1-Wechsel in der Folge g_0(p_0), \ldots, g_0(p_7), g_0(p_0)
                                                                                                                            (b)
          & g_0(p_0) \cdot g_0(p_2) \cdot g_0(p_6) = 0
                                                                                                                           (c)
          \& g_0(p_0) \cdot g_0(p_4) \cdot g_0(p_6) = 0 )
                                                                                                                           (d)
        g_1(p) = 0;
      else
        g_1(p) = g_0(p);
   for (p \in \mathbb{Z}^2) {
                                                                                                             /* 2. Schritt */
      if (g_1(p) = 1
          \& 2 \le \sum_{i=0}^{7} g_1(p_i) \le 6
                                                                                                                           (a)
          & 1 = Anzahl der 0-1-Wechsel in der Folge g_1(p_0), \ldots, g_1(p_7), g_1(p_0)
                                                                                                                           (b)
                                                                                                                           (c')
          \& g_1(p_0) \cdot g_1(p_2) \cdot g_1(p_4) = 0
                                                                                                                           (d')
          \& g_1(p_2) \cdot g_1(p_4) \cdot g_1(p_6) = 0 )
        g_2(p) = 0;
      else
        g_2(p) = g_1(p);
   }
}
while g_2 \neq g_0;
gib S = g_2^{-1}(1) aus.
```

Algorithmus 6.3: Verdünnungsalgorithmus

Bemerkungen (zum Algorithmus 6.3).

- 1) Zum Schluss ist $S := g_2^{-1}(1)$ ist eine skelettähnliche Menge für K.
- 2) (a) bedeutet, daß p zwischen 2 und 6 Nachbarn im momentanen Skelett $S = g^{-1}(1)$ hat; damit werden die Bedingungen 1) und 3) aus 6.5.1 erfüllt.
- 3) (c) und (d) bedeuten $(p_0 \notin S)$ oder $(p_6 \notin S)$ oder $(p_2 \notin S$ und $p_4 \notin S)$, d.h. p ist ein östlicher oder südlicher Randpunkt oder ein nordwestlicher Eckpunkt (denn $p_3 \notin S$ wegen (b)). (c') und (d') bedeuten $(p_2 \notin S)$ oder $(p_4 \notin S)$ oder $(p_0 \notin S$ und $p_6 \notin S)$, d.h. p ist ein nördlicher oder westlicher Randpunkt oder ein südöstlicher Eckpunkt (denn $p_7 \notin S$ wegen (b)).
- 4) Dadurch, daß zwei Schritte nacheinander ausgeführt werden, in denen Pixel an unterschiedlich geneigten Stellen des Randes entfernt werden, wird vermieden, daß Wege unterbrochen werden.
- 5) In manchen, allerdings uninteresanten Fällen liefert der Algorithmus nicht das Gewünschte:

6.6 Pixelweise logische und morphologische Operationen

 K_1 und K_2 seien zwei Binärobjekte. Definiere:

 $K_1 \text{ AND } K_2 := K_1 \cap K_2, \quad K_1 \text{ OR } K_2 := K_1 \cup K_2 \text{ und } K_1 \text{ EXOR } K_2 := (K_1 \cup K_2) \setminus (K_1 \cap K_2).$

Morphologische Operationen hatten wir schon in 2.8 eingeführt. In Verbindung mit logischen Operationen ergeben sich weitere nützliche Operatoren, z.B. der *Custer-Operator* (nach G.H.A. Custer)

 $\operatorname{Custer}(K) := \operatorname{Dilatation}(K) EXOR \operatorname{Erosion}(K).$

 $\operatorname{Custer}(K)$ ergibt einen Saum entlang des Randes von K. Vergleiche Anwendungsbeispiel im Anhang A.1.

6.7 Double thresholding und feature-AND

Mit Feature sind hier Zusammenhangskomponenten von Binärobjekten gemeint. Gegeben seien zwei Binärobjekte K_1 und K_2 (also zwei Binärbilder, die K_1 bzw. K_2 z.B. als schwarze Objekte auf weißem Hintergrund zeigen). Man definiert nun ein neues Binärobjekt

 K_1 feature-AND $K_2 :=$ Vereinigung aller Zusammenhangskomponenten Z von K_2 , für die $Z \cap K_1 \neq \emptyset$. Bemerkung. Diese Operation ist nicht kommutativ! Sie ist keine pixelweise AND-Verknüpfung. Beispiel.

+ 1 0 2 1023 1 0 2 1 0 2 K_1 K_2 K_1 feature-AND K_2 K_2 feature-AND K_1

Die Anwendung von feature-AND ist insbesondere im Verbund mit doppelten Schwellwertentscheidungen sinnvoll.

Typische Situation: Ein Grauwertbild wird mit zwei Schwellen ϑ_1 und ϑ_2 binärisiert; K_1 und K_2 seien die entstehenden Binärobjekte. K_2 enthalte die interessierenden Strukturen sehr deutlich, allerdings auch zusätzliche Komponenten, die als Störungen eingestuft werden. K_1 enthalte keine Störungen; die interessierenden Strukturen seien aber nur unvollständig mit Lücken zu sehen.

Dann enthält K_1 feature-AND K_2 genau die interessierenden Strukturen. Ähnlich arbeitet die bei Cannys Kantenfinder erwähnte Kantenverfolgung mit Hysterese. Vergleiche Anwendungsbeispiel im Anhang A.2.

6.8 Trennungsalgorithmen, der Wasserscheide-Algorithmus

Problem: Manchmal überlappen sich die interessierenden Bildstrukturen, so daß sie nach dem Binärisieren nicht disjunkte Zusammenhangskomponenten bilden. Sie müssen dann getrennt werden.

Eine Lösung dafür ist der Wasserscheide-Algorithmus, dessen Idee — wieder einmal — aus dem Kontinuierlichen stammt.

Sei $K \subset \mathbb{R}^2$ kompakt. Über K errichtet man ein Gebirge, nämlich den Graphen der *euklidischen Distanzabbildung* (*EDA*):

$$h: K \to [0, \infty[, h(p)] := \min\{\|p - q\|_2 \mid q \in \mathbb{R}^2 \setminus K\}.$$

Die Höhe des Gebirges im Punkt $p \in K$ ist also gerade gleich dem minimalen Abstand von p zum Komplement von K.

Auf dieses Gebirges läßt man nun Regen rieseln. Das Wasser fließe in negativer Gradientenrichtung. Markiert man alle Stellen, wo es aus mehr als einer Richtung hinfließt, so erhält man Trennungslinien zwischen den Gebirgskämen. Man nennt dieses Vorgehen *Wasserscheide-Algorithmus*. (Obwohl Wasserscheiden eigentlich eher die Stellen sind, von denen aus das Wasser in mehr als einer Richtung weiterfließen kann; wenn man -h statt h wählt, sind die Trennungslinien Wasserscheiden.)

6.8.1 Approximative Berechnung der EDA im Diskreten

Sei $K \subset \mathbb{Z}^2$ ein Binärobjekt und $h: K \to [0, \infty]$ die EDA.

Die direkte Berechnung von h gemäß Definition ist sehr ineffizient, weil man für jedes p das ganze Komplement von K durchsuchen muß. Daher wurden effizientere Verfahren entwickelt, die h aber nur approximativ berechnen.

Berechnung der EDA (Danielsson 1980)

- 1. Setze h(p) := 0 für $p \notin K$ und $h(p) := \infty$ für $p \in K$.
- 2. Durchlaufe das Bild zeilenweise von links nach rechts und von oben nach unten und tue an jeder Stelle p folgendes: Wenn $p \in K$, dann setze $h(p) := 1 + \min\{h(q) \mid q \in N_8(p), q \neq p\}$.
- 3. Wiederhole Schritt 2, aber diesmal von rechts nach links und von unten nach oben.

Beispiel



Bemerkung. Eine bessere Approximation der EDA erhält man, wenn man berücksichtigt, daß die Nachbarn in Diagonalrichtung $\sqrt{2}$ -fach weiter entfernt sind. Für den Wasserscheide-Algorithmus ist es aber einfacher, wenn die Werte der EDA nicht vom Typ *double*, sondern von einem möglichst kurzen *int*-Typ sind. Daher verwendet man 1.5 statt $\sqrt{2}$ und berechnet 2*h*. D.h. in Schritt 2 und 3 führt man folgendes aus:

Wenn $p \in K$, dann setze $h(p) := \min\{h(q) + d(q, p) \mid q \in N_8(p)\}$, wobei d(q, p) = 2, wenn q und p horizontal oder vertikal benachbart sind, und d(q, p) = 3 sonst.

Bemerkung. Durch den maximalen Wert des verwendeten Datentyps wird die maximal zulässige Dicke von *K* begrenzt.

6.8.2 Der Wasserscheide-Algorithmus

 $K \subset \mathbb{Z}^2$ sei ein Binärobjekt, $h: K \to [\![0,N]\!]$ sei die EDA.

$$\begin{split} M &:= \emptyset \\ \text{for } (z = N; z \ge 0; z\text{--}) \\ \text{for } (p \in h^{-1}(z)) \\ &\text{if } (p \text{ berührt höchstens eine Zusammenhangskomponente von } M) \\ &M &:= M \cup \{p\}; \\ &\text{else} \\ & \text{markiere } p \text{ als einen Trennungspunkt;} \end{split}$$

Algorithmus 6.4: Wasserscheide-Algorithmus

Zum Schluß stimmt M mit K überein bis auf die Trennungslinien.

In praktischen Implementierungen stellt man M als ein Bildarray dar, in dem die Pixel von M Werte > 0 und die im Komplement den Wert 0 erhalten. Es ist zweckmäßig, den Pixeln in jeder Zusammenhangskomponente von M denselben Wert zu geben und diese Werte untereinander verschieden zu wählen. Dann läßt sich leicht feststellen, ob ein Pixel mehr als eine Zusammenhangskomponente berührt. Vergleiche Anwendungsbeispiel im Anhang A.3.

6.9 Beschreibung der Gestalt

Von zwei Mengen K_1, K_2 im \mathbb{R}^2 sagt man üblicherweise, daß sie dieselbe Gestalt haben, wenn sie sich durch eine Bewegung und/oder eine Größenänderung ineinander transformieren lassen. Im Diskreten sind der Übertragung dieser Definition wieder Grenzen gesetzt; insbesondere wird durch starke Verkleinerung die Gestalt bis zur Unkenntlichkeit verändert.

Um Binärobjekte bezüglich ihrer Gestalt unterscheiden zu können, versucht man numerische oder nicht numerische Größen zu finden, die unter Bewegungen und Größenänderungen invariant sind. Man unterscheidet Größen, die von der globalen Gestalt abhängen, sogenannte *globale Deskriptoren*, und solche, die nur von einem kleinen Stück der Kontur abhängen, sogenannte *lokale Deskriptoren*. Solche lokalen Deskriptoren können geometrisch bedeutsame Punkte des Randes sein, wie z.B. Ecken oder Symmetrien zwischen Konturstücken. Sie ermöglichen, auch teilweise verdeckte Objekte oft noch zu erkennen. Ihre stabile Bestimmung erfordert aber einigen Aufwand.

6.9.1 Globale Gestaltbeschreibung: Fourierdeskriptoren

Zur Erinnerung an Fourierreihen: Sei T > 0. Für $f \in L_1([0,T])$ und $n \in \mathbb{Z}$ heißt

$$\hat{f}_n = \frac{1}{T} \int_0^T f(t) e^{-2\pi i n \frac{t}{T}} dt$$
 der *n*-te Fourierkoeffizient von *f*.

Es gilt: $\lim_{n\to\infty} \hat{f}_n = 0$, und die Abbildung $\mathcal{F} : L_1([0,T]) \to c_0 =$ Menge der Nullfolgen, $f \mapsto (\hat{f}_n)$ ist injektiv und stetig.

Für $f \in L_1([0,T])$ heißt die formale, nicht unbedingt konvergente Reihe

$$\sum_{n \in \mathbb{Z}} \hat{f}_n e^{2\pi i n \frac{t}{T}} \quad \text{die Fourierreihe von } f.$$

Es ist $L_2([0,T]) \subset L_1([0,T])$. Die Folge der Funktionen $e^{2\pi i n \frac{t}{T}}$, $n \in \mathbb{Z}$, ist eine Orthonormalbasis in dem Hilbertraum $L_2([0,T])$. Es gilt $f = \sum_{n \in \mathbb{Z}} \hat{f}_n e^{2\pi i n \frac{t}{T}}$, wobei die Reihe bezüglich der L_2 -Norm konvergiert (das ist das Analogon zur Umkehrformel der Fouriertransformation auf \mathbb{R}^n).

Für $f \in L_2([0,T])$ ist $(\hat{f}_n)_{n \in \mathbb{Z}} \in l_2(\mathbb{Z}) = \{(a_n)_{n \in \mathbb{Z}} \mid \sum_n |a_n|^2 < \infty\}$. Die Abbildung $\mathcal{F} : L_2([0,T]) \to l_2(\mathbb{Z}), f \mapsto (\hat{f}_n)_n$ ist ein Hilbertraum-Isomorphismus.

Ist $f : \mathbb{R} \to \mathbb{C}$ lokal integrierbar und periodisch mit Periode T, so ist $f|_{[0,T]} \in L_1([0,T])$ und man bezeichnet die Fourierkoeffizienten von $f|_{[0,T]}$ als Fourierkoeffizienten von f.

6.9.1.1 Satz. Sei $f : \mathbb{R} \to \mathbb{C}$ periodisch mit Periode T > 0 und stetig.

- a) Ist f stückweise differenzierbar, so konvergiert die Fourierreihe gleichmäßig gegen f.
- b) Ist f r-mal stetig differencierbar, so gilt $\hat{f}_n = O(\frac{1}{n^r})$ für $n \to \infty$.

6.9.1.2 Grundidee der Beschreibung von Konturen durch Fourierdeskriptoren

Wir betrachten den kontinuierlichen Fall.

K sei eine einfach zusammenhängende, kompakte Teilmenge von \mathbb{R}^2 mit stückweise glattem Rand. Es gibt dann eine Parametrisierung $f : [0,T] \to \mathbb{R}^2$ des Randes nach der Bogenlänge (d.h. mit $||f'||_2 = 1$). Identifiziere \mathbb{R}^2 mit \mathbb{C} wie üblich und betrachte f als Abbildung nach \mathbb{C} . Sie wird eindeutig beschrieben durch die Folge ihrer Fourierkoeffizienten $(\hat{f}_n)_{n\in\mathbb{Z}}$.

Aus der Änderung der Fourierkoeffizienten bei einer Bewegung kann man auf die Bewegung zurückschließen. Man kann aber auch umgekehrt aus den Fourierkoeffizienten durch geeignete Normierungen sogenannte *Fourierdeskriptoren* gewinnen, die invariant unter Bewegungen und Größenänderungen sind und trotzdem noch die Gestalt der Kontur charakterisieren.

6.9. BESCHREIBUNG DER GESTALT

Man kann beispielsweise \hat{f}_0 weglassen, denn \hat{f}_0 ist der Schwerpunkt von K und $\frac{\hat{f}_n}{\hat{f}_1}$ ist invariant unter Größenänderungen. Literaturhinweis: [4], [8]

Fourierdeskriptoren nach Zahn und Roshies (siehe [12])

f sei eine Parametrisierung des Randes wie oben. $\vartheta : [0,T] \to \mathbb{R}$ gebe den Winkel zwischen der Tangente f'und der x-Achse (mit Vielfachheiten von 2π) an. Setze $\Theta(t) := \vartheta(\frac{t}{2\pi}T) - \vartheta(0) - t$ für $t \in [0,2\pi]$. $\hat{\Theta}_n$ seien die Fourierkoeffizienten von Θ . Sie ändern sich nicht, wenn K bewegt oder seine Größe verändert wird. Sie charakterisieren aber f bis auf Bewegungen und Skalierungsänderungen.

6.9.1.3 Nachteile

- Fourierdeskriptoren sind Parameter für die globale Gestalt; teilweise verdeckte Objekte können nicht beschrieben werden.
- Bei diskreten Bildern erhält man keine exakte Parametrisierung nach der Bogenlänge; die Bestimmung von ϑ ist ungenau. Empfindlich gegenüber Rauscheffekten.

Es gibt zahlreiche Ansätze zur lokalen Gestaltbeschreibung, die aber anders arbeiten.

6.9.2 Lokale Gestaltmerkmale

Idee: Finde geometrisch markante Punkte der Kontur und zerteile mit ihnen die Kontur in Stücke, die sich einfach beschreiben lassen, z.B. näherungsweise als Geraden- oder Kreisbogenstücke. Merkt man sich noch die relative Lage dieser Punkte zueinander, so hat man eine (größen- und bewegungsinvariante) Beschreibung der Gestalt. Man kann sie z.B. in Form eines attributierten Graphen abspeichern. Damit läuft die Erkennung von Objekten auf ein Graph-Matching hinaus. Und dies ist oft auch dann noch möglich, wenn das Objekt nur teilweise zu sehen ist, also der Graph nur teilweise bekannt ist.

Geometrisch markant bedeutet, daß die Punkte auch bei leichten Störungen und Verzerrungen der Objektkontur noch gut zu detektieren sind und sich ihre Positionen nicht sehr verändern.

Typische Kandidaten:

- Ecken, deren Winkel nicht zu flach sind.
- Punkte, in denen die Krümmung der Kontur deutliche, lokale Extrema hat.
- Wendepunkte, also Punkte, in denen die Krümmung der Kontur das Vorzeichen wechselt.

Alle diese Begriffe stammen (wie üblich) aus dem Kontinuierlichen; ihre Definitionen beinhalten Ableitungen bis zur 2. Ordnung einer Parametrisierung der Kontur. Deshalb kann eine direkte Übertragung ins Diskrete zu instabilen Ergebnissen führen.

Im Folgenden werden einige Algorithmen angegeben, die die direkte Übertragung differentieller, infinitesimaler Begriffe vermeiden und deshalb recht robust gegenüber kleinen Störungen sind. Sie sind nicht sehr kompliziert, und ihre Berechnung lässt sich massiv parallel durchführen.

Im Nachfolgenden sei $K \subset \mathbb{Z}^2$ ein Binärobjekt, ∂K der Rand von K (bezüglich der 4- oder 8-Nachbarschaft), $D(p,r) = \{q \in \mathbb{Z}^2 \mid ||q-p||_2 \leq r\}$ eine Art Kreisscheibe um p mit Radius r, und $g_{\sigma}(t) = e^{-t^2/2\sigma^2}$ mit $\sigma > 0$.

6.9.2.1 Tangenten und Normalen

Für $p \in \partial K$ sei

$$u_p := \sum_{q \in D(p, 2\sigma) \cap K} g_\sigma(\|q - p\|_2) \Big(q - p\Big)$$

Dieser Vektor steht meist nahezu senkrecht auf dem Rand bei p. Durch Normierung erhält man mit $\nu_p := \frac{u_p}{\|u_p\|_2}$ eine gute Näherung für den nach innen zeigenden Normalenvektor in p.



"Physikalische" Interpretation: Belege $D(p, 2\sigma) \cap K$ mit einer Massenverteilung der Dichte $q \mapsto g_{\sigma}(||q-p||_2)$ und berechne deren Schwerpunkt s_p . Dann gilt

$$u_p = (s_p - p) \sum_{q \in D(p, 2\sigma) \cap K} g_\sigma(||q - p||_2) \text{ und somit } \nu_p = \frac{s_p - p}{||s_p - p||}.$$

Probleme:

- Wenn das Objekt K bei p sehr dünn ist, so daß in der Summe fast nur nahezu tangentielle Anteile 1. auftreten, die sich gegenseitig kompensieren, dann kann $\|u_p\|_2$ sehr klein werden.
- 2.Wenn die Kontur stark gekrümmt ist, können in der Kreisscheibe um p mit Radius 2σ Teile von K liegen, die mit der lokalen Gestalt der Kontur bei p nichts zu tun haben. Dadurch wird die Richtung von u_p verfälscht. Dem kann man begegnen durch folgende

Verbesserung: Erstrecke die Summe nur über die Zusammenhangskomponente von $D(p, 2\sigma) \cap K$, die p enthält.

Umfangschätzung: Mit Hilfe der Normalen kann man eine brauchbare Näherung für den Umfang berechnen. Sie geht von der üblichen Formel für die Länge des Graphen einer differenzierbaren Funktion aus; sie verwendet jedoch keinerlei serielle Konturverfolgung und ist massiv parallel berechenbar.

 $\operatorname{Rand}_4(K) := \{p \in K \mid N_4(p) \cap (\mathbb{Z}^2 \setminus K) \neq \emptyset\}$ sei die Menge der Randpixel bezüglich der 4-Nachbarschaft. Und für $p \in \text{Rand}_4(K)$ sei $\nu_p = (\nu_{p,x}, \nu_{p,y})$ ein Normalenvektor an den Rand in p. Setze nun

$$f(p) := \begin{cases} \sqrt{1 + (\frac{\nu_{p,y}}{\nu_{p,x}})^2} & \text{, falls } |\nu_{p,x}| \ge |\nu_{p,y}| \\ \sqrt{1 + (\frac{\nu_{p,y}}{\nu_{p,y}})^2} & \text{, sonst} \end{cases}$$

Dann ist $U = \sum_{p \in \text{Rand}_4(K)} f(p)$ eine brauchbare Näherung für den Umfang.

6.9.2.2 Krümmung

Für $p \in \partial K$ sei

$$r_p := \frac{1}{\alpha_p} \sum_{q \in D(p, 2\sigma) \cap \partial K} g_{\sigma}(\|q - p\|_2) \cdot q$$

mit $\alpha_p = \sum_{q \in D(p,2\sigma) \cap \partial K} g_{\sigma}(||q-p||_2).$ r_p ist der Schwerpunkt der Masseverteilung auf $D(p,2\sigma) \cap \partial K$, die die Dichte $q \mapsto g_{\sigma}(||q-p||_2)$ hat.

Setze $k_p = r_p - p$. Der Wert von $||k_p||_2$ ist nicht gleich der Absolutkrümmung, hängt aber monoton von ihr ab. Er kann daher für qualitative Entscheidungen über die Absolutkrümmung der Kontur in p herangezogen werden.

Das Vorzeichen der Krümmung, also ob die Kontur bei p konvex oder konkav ist, kann man daran ablesen, ob k_p und ν_p in den selben Halbraum zeigen oder in entgegengesetzte, also

> Kontur bei p konvex $\Leftrightarrow \langle k_p, \nu_p \rangle > 0$ Kontur bei p konkav $\Leftrightarrow \langle k_p, \nu_p \rangle < 0$

Setze daher $\kappa(p) := \langle k_p, \nu_p \rangle$ als Schätzung für die Krümmung. Verbesserungen:

Erstrecke die Summe bei der Berechnung von r_p nur über die q aus derjenigen Zusammenhangskomponente 1. Z(p) von $D(p, 2\sigma) \cap \partial K$, die p enthält.







6.9. BESCHREIBUNG DER GESTALT

2. Bei geradlinigen Konturstücken ist r_p infolge der diskreten Pixelstruktur nicht exakt gleich p und liegt mal diesseits und mal jenseits der Kontur, so daß ρ_p zwischen betraglich kleinen positiven und negativen Werten schwankt.

Dies kann man verhindern, indem man das Vektorfeld r_p lokal ein wenig glättet; im einfachsten Fall durch ein lineares Gaußfilter, besser jedoch durch ein nichtlineares Gaußfilter: Ersetze r_p durch

$$\frac{1}{\alpha_p} \sum_{q \in Z(p)} g_{\sigma}(\|q-p\|_2) g_{\tau}(\|r_p-r_q\|_2) \cdot r_q \quad \text{mit} \quad \alpha_p = \sum_{q \in Z} g_{\sigma}(\|q-p\|_2) g_{\tau}(\|r_q-r_p\|_2).$$

6.9.2.3 Wendepunkte

Stellen p, in denen κ das Vorzeichen wechselt, wenn man bei p der Kontur entlang läuft.

6.9.2.4 Robuste Bestimmung lokaler Extrema

 $f: R \to \mathbb{R}$ Function, h > 0, b > 0, A =Kreiszylinder mit Radius b und Höhe h, (p, f(p)) Grundpunkt.



f hat in p lokales Maximum \Leftrightarrow Die Zusammenhangskomponente Z von graph $(f) \cap A$, die p enthält, trifft die Wände und den oberen Deckel von A nicht (außer in (p, f(p))). Lokale Minima analog.

6.9.2.5 Lokale Krümmungsextrema

Ersetzt man in 6.9.2.4 die Funktion f durch die nur auf ∂K definierte Krümmungsfunktion κ , so kann man völlig analog lokale Krümmungsextrema von ∂K bestimmen.

6.9.2.6 Ecken und Enden

Ecken sind ausgeprägte Krümmungsextrema und können somit wie in 6.9.2.5 detektiert werden. Für Enden trifft dies nicht immer zu. So muß die Krümmung nicht unbedingt ein sehr deutliches Extremum annehmen, wenn das Objekt dick ist. Und bei dünnen, linienhaften Objekten kann man aufgrund der Krümmung nicht zwischen Ecken und Enden unterscheiden. Manchmal ist die Unterscheidung auch nicht klar zu treffen. Statt mit lokalen Krümmungsextrema kann man auch ähnlich wie in 6.9.2.4 bei der Bestimmung lokaler Extrema von Funktionen mit rechteckigen Schablonen Ecken und Enden finden.



KAPITEL 6. VERARBEITUNG VON BINÄRBILDERN

Für $p \in \partial K$ sei ν_p die äußere Normale. A sei ein Rechteck wie in der Zeichnung. Z sei diejenige Zusammenhangskomponente von $A \cap K$, die p enthält. Dann wird p als Ecke markiert, wenn die folgenden Eigenschaften erfüllt sind:

- 1) Z trifft die Seitenwände nicht.
- 2) Z trifft außer in p den Boden nicht.

Mit b und h kann man die Form der Ecken angeben, die markiert werden sollen. Bei abgerundeten Ecken und breitem, niedrigem Rechteck A, d.h. b groß und h klein, werden mehrere nebeneinanderliegende Punkte markiert.

Dieses Verfahren eignet sich auch zum Finden von Ecken linienhafter Objekte. Allerdings liefert die Normalenschätzung nach 6.9.2.1 bei geraden, dünnen Objekten unzuverlässige Werte, was man aber daran erkennen kann, daß der Schwerpunktsvektor u_p recht klein ist. Schließt man solche Stellen von der Markierung von vornherein aus, so kann man Ecken und Enden recht sicher markieren. Wenn man b klein wählt, werden nicht allzu spitze Ecken nicht markiert.

Beispiele.





Kapitel 7

Beschreibung linearer Filter im Frequenzbereich

In diesem Kapitel bezeichnet $i = \sqrt{-1}$ stets die imaginäre Einheit.

Wir werden oft parallel den kontinuierlichen und den diskreten Fall betrachten, wobei wir im Diskreten nur den endlichen Fall untersuchen.

 $L_1(\mathbb{R}^n)$ = Banachraum der Lebesgue-integrierbaren Funktionen $f: \mathbb{R}^n \to \mathbb{C}$.

 $\mathcal{C}_0(\mathbb{R}^n) =$ Banachraum der stetigen, bei ∞ verschwindenden Funktionen $f : \mathbb{R}^n \to \mathbb{C}$ mit der Supremumsnorm $\|f\|_{\infty} = \sup_{p \in \mathbb{R}^n} |f(p)|.$

7.1 Die Fouriertransformation

7.1.1 Die Fouriertransformation auf \mathbb{R}^n

Für $f \in L_1(\mathbb{R}^n)$ und $\omega \in \mathbb{R}^n$ sei

$$\hat{f}(\omega) = \int_{\mathbb{R}^n} f(p) \, e^{-2\pi i \langle \omega, p \rangle} \, dp.$$

Es gilt:

a) $\hat{f} \in \mathcal{C}_0(\mathbb{R}^n)$. (Lemma von Riemann-Lebesgue)

b) $\mathcal{F}: L_1(\mathbb{R}^n) \to \mathcal{C}_0(\mathbb{R}^n), f \mapsto \hat{f}$ ist linear, stetig und injektiv.

 \hat{f} heißt die Fouriertransformierte von f, und \mathcal{F} heißt die Fouriertransformation (FT). Der Fall n = 1: $\hat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-2\pi i \omega t} dt$. Der Fall n = 2: $\hat{f}(\omega_x, \omega_y) = \int_{\mathbb{R}^2} f(x, y) e^{-2\pi i (\omega_x x + \omega_y y)} dx dy$. **Umkehrformel**: Sind $f \in L_1(\mathbb{R}^n)$ und $\hat{f} \in L_1(\mathbb{R}^n)$, so gilt

$$f(p) = \int_{\mathbb{R}^n} \hat{f}(\omega) e^{2\pi i \langle p, \omega \rangle} d\omega$$
 fast-überall auf \mathbb{R}^n .

Interpretation (für n = 1): $f(t) = \int_{\mathbb{R}} \hat{f}(\omega) e^{2\pi i \omega t} d\omega$.

f ist dargestellt als Überlagerung von kontinuierlich vielen harmonischen Schwingungen der Gestalt $t \mapsto e^{2\pi i\omega t}$ mit der Frequenz $\omega \in \mathbb{R}$. Setze $\hat{f}(\omega) = A(\omega) e^{i\varphi(\omega)}$ mit $A(\omega) \ge 0$. Dann ist $A(\omega)$ die Amplitude der Schwingung mit Frequenz ω und $\varphi(\omega)$ ihre Phasenverschiebung. **Bemerkung.** Die Umkehrformel ist *keine* Entwicklung von f nach der Schauderbasis bestehend aus den Schwingungen $e^{2\pi i\omega t}$ mit $\omega \in \mathbb{R}$.

 $f \mapsto \check{f}$ mit $\check{f}(\omega) = \int_{\mathbb{R}} f(p) e^{2\pi i \langle p, \omega \rangle} dp$ heißt die *inverse Fouriertransformation*. Beispiele.

- 1) Sei $f(t) = \operatorname{rect}_a(t) := \begin{cases} 1 & \text{, für } |t| < \frac{a}{2} \\ 0 & \text{, sonst} \end{cases}$. Dann gilt: $\hat{f}(\omega) = \frac{\sin(a\pi\omega)}{\pi\omega}$.
- 2) Sei $f(t) = e^{-at^2}$. Dann gilt: $\hat{f}(\omega) = \sqrt{\frac{\pi}{a}} e^{-\frac{\pi^2 \omega^2}{a}}$.

Insbesondere gilt für die Gaußfunktion $g_{\sigma}(t) = e^{-\frac{t^2}{2\sigma^2}}$, daß $\hat{g_{\sigma}}(\omega) = \sqrt{2\pi\sigma} e^{-2\pi^2 \sigma^2 \omega^2} = \sqrt{2\pi\sigma} g_{\alpha}(\omega)$ mit $\alpha = \frac{1}{2\pi\sigma}$, also $\sigma \cdot \alpha = \frac{1}{2\pi}$.

3) $\widehat{\frac{\partial f}{\partial x_j}} = 2\pi i \omega_j \cdot \hat{f}, \ \widehat{\Delta f} = -(2\pi)^2 \|\omega\|_2^2 \hat{f} \text{ für } f \in \mathcal{C}^1(\mathbb{R}^n).$

4)
$$\hat{f} = f_{-} \text{ mit } f_{-}(t) = f(-t).$$

7.1.2 Die diskrete Fouriertransformation

Der eindimensionale Fall:

Sei $N \in \mathbb{N}, \mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ die Quotientengruppe. Für jede Funktion $f : \mathbb{Z}_N \to \mathbb{C}$ und jedes $\omega \in \mathbb{Z}_N$ sei

$$\hat{f}(\omega) = \sum_{p \in \mathbb{Z}_N} f(p) e^{-\frac{2\pi i}{N}\omega p}.$$

Beachte: $\exp(-\frac{2\pi i}{N}\omega p)$ wird dadurch definiert, daß man statt ω und p Repräsentanten aus \mathbb{Z} einsetzt. Man überlege sich, daß es keine Rolle spielt, welche man auswählt.

Meist repräsentiert man die Elemente von \mathbb{Z}_N durch ganze Zahlen vermöge der kanonischen Quotientenabbildung $\mathbb{Z} \to \mathbb{Z}_N$, indem man diese auf ein Intervall der Länge N einschränkt, wodurch sie bijektiv wird. Üblicherweise wählt man das Intervall $[\![0, N-1]\!]$ als Definitionsbereich von f und meist auch von \hat{f} .

Obige Definition erhält dann folgende Gestalt:

Für jede Funktion $f : [0, N-1] \to \mathbb{C}$ und jedes $k \in [0, N-1]$ sei

$$\hat{f}(k) = \sum_{j=0}^{N-1} f(j) e^{-\frac{2\pi i}{N}kj}.$$

Beachte: Die Funktionen $f, \hat{f} : [0, N - 1]] \to \mathbb{C}$ kann man als Vektoren $(f(0), \ldots, f(N - 1)) \in \mathbb{C}^N$ und $(\hat{f}(0), \ldots, \hat{f}(N - 1)) \in \mathbb{C}^N$ auffassen. Dann erhält man eine Abbildung

$$\mathcal{F}: \mathbb{C}^N \to \mathbb{C}^N, \ f \mapsto \hat{f}.$$

 \hat{f} heißt die diskrete Fouriertransformierte von f und \mathcal{F} die diskrete Fouriertransformation (DFT) der Länge N. Es gilt: \mathcal{F} ist linear und bijektiv, also ein Vektorraumautomorphismus.

Der zweidimensionale Fall:

Seien $N_x, N_y \in \mathbb{N}$. Für $f : [0, N_x - 1] \times [0, N_y - 1] \rightarrow \mathbb{C}$ und $k_x \in [0, N_x - 1], k_y \in [0, N_y - 1]$ sei

$$\hat{f}(k_x, k_y) = \sum_{x=0}^{N_x - 1} \sum_{y=0}^{N_y - 1} f(x, y) e^{-\frac{2\pi i}{N_x} k_x x} e^{-\frac{2\pi i}{N_y} k_y y}.$$

 \hat{f} heißt die diskrete Fouriertransformierte von f und $\mathcal{F} : \mathbb{C}^{N_x \times N_y} \to \mathbb{C}^{N_x \times N_y}$, $f \mapsto \hat{f}$ die diskrete Fouriertransformation auf $[\![0, N_x - 1]\!] \times [\![0, N_y - 1]\!]$. Es gilt: \mathcal{F} ist ein Automorphismus. Bemerke: Für $N_y = 1$ erhält man den eindimensionalen Fall.

7.1. DIE FOURIERTRANSFORMATION

Die inverse DFT, die Umkehrformel:

Für jedes $f: [\![0, N_x-1]\!]\times [\![0, N_y-1]\!] \to \mathbb{C}$ gilt

$$f(x,y) = \frac{1}{N_x N_y} \sum_{k_x=0}^{N_x - 1} \sum_{k_y=0}^{N_y - 1} \hat{f}(k_x, k_y) e^{\frac{2\pi i}{N_x} k_x x} e^{\frac{2\pi i}{N_y} k_y y}.$$

7.1.3 Die DFT als Koordinatenwechsel

Sei $R = [\![0, N_x - 1]\!] \times [\![0, N_y - 1]\!].$ Für $(u, v) \in R \subset \mathbb{Z}^2$ definieren wir

$$\delta_{(u,v)} : R \to \mathbb{C} \text{ durch } \delta_{(u,v)}(x,y) = \begin{cases} 1 & \text{, falls } (x,y) = (u,v) \\ 0 & \text{, sonst} \end{cases}$$

und

$$E_{(u,v)}: R \to \mathbb{C} \text{ durch } E_{(u,v)}(x,y) = \exp\left(2\pi i \left\langle \left(\frac{u}{N_x}, \frac{v}{N_y}\right), (x,y) \right\rangle \right) = e^{\frac{2\pi i}{N_x}ux} e^{\frac{2\pi i}{N_y}vy}.$$

Stellt man sich Abbildungen $f : R \to \mathbb{C}$ als Vektoren im \mathbb{C}^R vor, so sind die $\delta_{(u,v)}$ mit $(u,v) \in R$ genau die kanonischen Einheitsvektoren im \mathbb{C}^R ; die $\delta_{(u,v)}$ mit $(u,v) \in R$ bilden also eine Orthonormalbasis des \mathbb{C}^R .

7.1.3.1 Lemma.

- a) $k_x l_x \in N_x \mathbb{Z}$ und $k_y l_y \in N_y \mathbb{Z} \Leftrightarrow E_{(k_x, k_y)} = E_{(l_x, l_y)}$.
- b) Ist $k_x l_x \notin N_x \mathbb{Z}$ oder $k_y l_y \notin N_y \mathbb{Z}$, so sind $E_{(k_x,k_y)}$ und $E_{(l_x,l_y)}$ orthogonal bezüglich des euklidischen Skalarprodukts im \mathbb{C}^R .

c)
$$||E_{(k_x,k_y)}||_2^2 = N_x N_y.$$

Beweis.

- a) Der erste Teil ist klar, weil $e^{2\pi i m} = 1$ für alle $m \in \mathbb{Z}$.
- b) Sei o.E. $k_x l_x \notin N_x \mathbb{Z}$. Setze $W_x = e^{\frac{2\pi i}{N_x}}$ und $W_y = e^{\frac{2\pi i}{N_y}}$.

$$\begin{split} \left\langle E_{(k_x,k_y)}, E_{(l_x,l_y)} \right\rangle &= \sum_{(x,y)\in R} W_x^{k_x x} W_y^{k_y y} W_x^{-l_x x} W_y^{-l_y y} = \sum_y W_y^{y(k_y-l_y)} \sum_x W_x^{x(k_x-l_x)} \\ &= \sum_y W_y^{y(k_y-l_y)} \frac{1 - W_x^{N_x(k_x-l_x)}}{1 - W_x^{k_x-l_x}}, \, \mathrm{da} \; W_x^{k_x-l_x} \neq 1 \\ &= 0, \, \mathrm{da} \; W_x^{N_x(k_x-l_x)} = 1. \end{split}$$

c) Nachrechnen.

7.1.3.2 Korollar. Die Familie $\left(\frac{1}{\sqrt{N_x N_y}} E_{(k_x, k_y)} \mid (k_x, k_y) \in R\right)$ ist eine Orthonormalbasis des \mathbb{C}^R .

7.1.3.3 Bemerkung. Es gelten:

$$\exp\left(\frac{2\pi i}{N_x}k_x\right) = \exp\left(\frac{2\pi i}{N_x}(k_x - N_x)\right) \text{ und } \exp\left(\frac{2\pi i}{N_y}k_y\right) = \exp\left(\frac{2\pi i}{N_y}(k_y - N_y)\right)$$

und somit für beliebige $(k_x, k_y), (l_x, l_y) \in \mathbb{Z}^2$:

$$E_{(k_x,l_y)} = E_{(k_x - N_x,l_y)}$$
 und $E_{(l_x,k_y)} = E_{(l_x,k_y - N_y)}$

Seien nun N_x und N_y geradzahlig.

Dann sind

$$\begin{bmatrix} \frac{N_x}{2}, N_x - 1 \end{bmatrix} \to \begin{bmatrix} -\frac{N_x}{2}, -1 \end{bmatrix}, \ k_x \mapsto k_x - N_x \text{ und } \begin{bmatrix} \frac{N_y}{2}, N_y - 1 \end{bmatrix} \to \begin{bmatrix} -\frac{N_y}{2}, -1 \end{bmatrix}, \ k_y \mapsto k_y - N_y = N_y =$$

Bijektionen.

Somit ist

$$\left(E_{(k_x,k_y)} \mid (k_x,k_y) \in [\![-\frac{N_x}{2},\frac{N_x}{2}-1]\!] \times [\![-\frac{N_y}{2},\frac{N_y}{2}-1]\!]\right)$$

dieselbe Orthogonalbasis wie

$$(E_{(k_x,k_y)} \mid (k_x,k_y) \in [[0,N_x-1]] \times [[0,N_y-1]]);$$

lediglich der Indexbereich ist verschoben.

Diesen verschobenen Indexbereich benutzt man häufig, weil dadurch die Analogie zur Fouriertransformation auf $\mathbb R$ deutlicher wird.

Die Asymmetrie der Intervalle $\left(-\frac{N_x}{2} \text{ und } -\frac{N_y}{2} \text{ gehören dazu, } \frac{N_x}{2} \text{ und } \frac{N_y}{2} \text{ nicht}\right)$ ist bei genauerer Betrachtung keine Asymmetrie; denn, weil man eigentlich in \mathbb{Z}_{N_x} bzw. \mathbb{Z}_{N_y} ist, muß man modulo N_x bzw. N_y rechnen, und dann gilt $-\frac{N_x}{2} \equiv \frac{N_x}{2} \mod N_x$ und $-\frac{N_y}{2} \equiv \frac{N_y}{2} \mod N_y$, das heißt, man muß die Intervalle als Diskretisierung der Einheitskreislinie auffassen vermöge $k \mapsto e^{\frac{2\pi i}{N}k}$.



7.1.3.4 Bemerkung. Wählt man $R = [0, N_x - 1] \times [0, N_y - 1]$ als Repräsentantensystem von $\mathbb{Z}_{N_x} \times \mathbb{Z}_{N_y}$, so kann man die DFT als Automorphismus $\mathcal{F} : \mathbb{C}^R \to \mathbb{C}^R$, $f \mapsto \hat{f}$ auffassen. \mathcal{F} ist bis auf den Faktor $\sqrt{N_x N_y}$ lediglich ein *Koordinatenwechsel* von der Orthonormalbasis $(\delta_{(u,v)} \mid (u,v) \in R)$ zu der Orthonormalbasis $(\frac{1}{\sqrt{N_x N_y}} E_{(u,v)} \mid (u,v) \in R)$. Die Werte von $\hat{f}/\sqrt{N_x N_y}$ sind die Koeffizienten der Entwicklung von f nach dieser Basis; das besagt gerade der Umkehrsatz.

Bemerkung. Im Kontinuierlichen stimmt das nicht! Es liegt eine Verallgemeinerung des Begriffs einer Basisentwicklung vor.

Statt der komplexwertigen Funktionen $E_{(k_x,k_y)}$ kann man auch folgende reellwertige Funktionen als Basis wählen:

$$\frac{1}{2} \left(E_{(k_x,k_y)} + E_{(-k_x,-k_y)} \right) (x,y) = \cos \left(2\pi \left\langle \left(\frac{x}{N_x}, \frac{y}{N_y} \right), (k_x,k_y) \right\rangle \right)$$

and
$$\frac{1}{2i} \left(E_{(k_x,k_y)} - E_{(-k_x,-k_y)} \right) (x,y) = \sin \left(2\pi \left\langle \left(\frac{x}{N_x}, \frac{y}{N_y} \right), (k_x,k_y) \right\rangle \right)$$

Das sind Streifen senkrecht zu (k_x, k_y) mit k_x Perioden in x-Richtung und k_y Perioden in y-Richtung. Dabei sind $0 \le k_x \le \frac{N_x}{2}$ und $0 \le k_y \le \frac{N_y}{2}$.

7.1.3.5 Lemma.

a) f reellwertig $\Leftrightarrow \hat{f}(-k_x, -k_y) = \overline{\hat{f}(k_x, k_y)}$ für alle (k_x, k_y) .

- b) Für $(u, v) \in \mathbb{Z}^2$ sei $\tau_{(u,v)} f(x, y) = f(x u, y v)$ (Argumente modulo N_x bzw. N_y). Dann gilt: $(\widehat{\tau_{(u,v)}f})(k_x, k_y) = E_{(-u,-v)}(k_x, k_y)\widehat{f}(k_x, k_y)$ für alle (k_x, k_y) .
- c) $\langle \hat{f}, \hat{g} \rangle = N_x N_y \langle f, g \rangle.$
- d) $\|\hat{f}\|_2 = \sqrt{N_x N_y} \|f\|_2.$
- e) $\hat{\delta}_{(u,v)} = E_{(-u,-v)}$ und $\hat{\delta}_{(0,0)} = 1$.
- f) $\widehat{E_{(u,v)}} = N_x N_y \,\delta_{(u,v)}$ und $\hat{1} = N_x N_y \,\delta_{(0,0)}$.

64

7.1. DIE FOURIERTRANSFORMATION

7.1.4 Interpretation, Sprechweisen, Konventionen

Sei $f : R \to \mathbb{C}$ eine Funktion mit $R = [[0, N_x - 1]] \times [[0, N_y - 1]]$. Wir setzen von nun an N_x und N_y stets als geradzahlig voraus.

Als Darstellung von f im Ortsbereich bezeichnet man f selbst oder, wenn man so will, die Entwicklungskoeffizienten von f nach der Orthonormalbasis $(\delta_{(u,v)} | (u,v) \in R)$.

Als Darstellung von f im Frequenzbereich bezeichnet man die diskrete Fouriertransformierte \hat{f} von f, die man, wie wir in 7.1.3 gesehen haben, bis auf den Faktor $\sqrt{N_x N_y}$ als die Entwicklungskoeffizienten von f nach der

Orthonormalbasis $\left(\frac{1}{\sqrt{N_x N_y}} E_{(u,v)} \mid (u,v) \in R\right)$ auffassen kann. Der Definitionsbereich von \hat{f} ist mathematisch gesehen $\mathbb{Z}_{N_x} \times \mathbb{Z}_{N_y}$. Aber in praktischen Rechnungen muß man Repräsentantensysteme verwenden. Rein rechentechnisch wählt man meist $[0, N_x - 1] \times [0, N_y - 1]$, z.B. wenn man \hat{f} als zweidimensionales Array in C darstellt (in FFT-Routinen, siehe später).

Für die anschauliche Interpretation ist es günstiger, $\left[\!\left[-\frac{N_x}{2}, \frac{N_x}{2}\right]\!\right] \times \left[\!\left[-\frac{N_y}{2}, \frac{N_y}{2}\right]\!\right]$ zu wählen.

 \hat{f} wird auch das *Spektrum* von f genannt; $|\hat{f}|$ heißt das *Amplitudenspektrum* und $\frac{\hat{f}}{|\hat{f}|}$ das *Phasenspektrum* (nicht definiert bei Nullstellen von \hat{f}).

Statt $|\hat{f}|$ stellt man auch oft $\log |\hat{f}|$ oder $\log(1+|\hat{f}|)$ dar. $|\hat{f}|^2$ heißt das Leistungsspektrum.

Die Punkte (k_x, k_y) im Definitionsbereich von \hat{f} werden Ortsfrequenzen oder Wellenzahlen genannt.

Vereinbarung (für die Übungen). Ein komplexes Signal $f : [0, N_x - 1]] \times [0, N_y - 1]] \rightarrow \mathbb{C}$ wird wie ein Grauwertbild als eindimensionales Array dargestellt (mit derselben Elementreihenfolge), nur hat jedes Element den Datentyp *COMPLEX*, der folgendermaßen definiert ist:

typedef struct{double r,i;} complex;
#define COMPLEX complex;

Bei Bedarf kann so ein Array auch als zweidimensionales Array adressiert werden (wie bei Grauwertbildern).

Die bereitgestellten Routinen zum Speichern und Lesen solcher Arrays machen dies byteweise. Dadurch sind diese Dateien plattformabhängig! Denn die interne Darstellung von *double* kann variieren.

Die Dateien haben folgenden Header:

S5 /* Magic Number */ Nx Ny /* als ASCII-Strings */

Bei allen FFT-Routinen müssen ${\cal N}_x$ und ${\cal N}_y$ Zweier-Potenzen sein.

Wählt man $[\![-\frac{N_x}{2}, \frac{N_x}{2} - 1]\!] \times [\![-\frac{N_y}{2}, \frac{N_y}{2} - 1]\!]$ als Definitions bereich von \hat{f} , so ist \hat{f} folgendermaßen in einem zweidimensionalem Array abgelegt:

0,0	 $\frac{N_x}{2} - 1, 0$	$-\frac{N_x}{2}, 0$	 -1, 0
÷	÷	÷	:
$0, \frac{N_y}{2} - 1$	 $\frac{N_x}{2} - 1, \frac{N_y}{2} - 1$	$-\frac{N_x}{2}, \frac{N_y}{2}-1$	 $-1, \frac{N_y}{2} - 1$
$0, -\frac{N_y}{2}$	 $\frac{N_x}{2} - 1, -\frac{N_y}{2}$	$-\frac{N_x}{2},-\frac{N_y}{2}$	 $-1, -\frac{N_y}{2}$
:	÷	÷	:
0, -1	 $\frac{N_x}{2} - 1, -1$	$-\frac{N_x}{2}, -1$	 -1, -1

$-\frac{N_x}{2}, -\frac{N_y}{2}$	•••	$-1, -\frac{N_y}{2}$	$0, -\frac{N_y}{2}$	 $\frac{N_x}{2} - 1, -\frac{N_y}{2}$
:			••••	
$-\frac{N_x}{2}, -1$	•••	-1, -1	0, -1	 $\frac{N_x}{2} - 1, -1$
$-\frac{N_x}{2}, 0$		-1, 0	0, 0	 $rac{N_x}{2}-1, 0$
:				:
$-\frac{N_x}{2}, \frac{N_y}{2}-1$		$-1, \frac{N_y}{2} - 1$	$0, \frac{N_y}{2} - 1$	 $\frac{N_x}{2} - 1, \frac{N_y}{2} - 1$

Die Programme S5show* rollen das Array vor der Darstellung zyklisch um $\frac{N_x}{2}$ bzw. $\frac{N_y}{2}$ Plätze schräg nach unten rechts, so daß sich folgende Anordnung ergibt:

7.1.5 Schnelle DFT-Algorithmen, FFT

Wegen $\hat{f}(k_x, k_y) = \sum_{x=0}^{N_x-1} e^{-\frac{2\pi i}{N_x}k_x x} \left(\sum_{y=0}^{N_y-1} e^{-\frac{2\pi i}{N_y}k_y y} f(x, y) \right)$ kann die zweidimensionale DFT als Komposition zweier eindimensionaler DFT betrachtet werden. Zur Berechnung der eindimensionalen DFT gibt es sogenannte *FFT-Algorithmen (fast fourier transform)*, die geschickt die Funktionalgleichung der Exponentialfuntion ausnutzen, um die Anzahl der Rechenoperationen ganz wesentlich zu senken: Statt $\Theta(n^2)$ Additionen und Multiplikationen sind es nur $\Theta(n \log(n))$.

Allerdings muß man im Zweidimensionalen sehr aufpassen, daß man den Geschwindigkeitsvorteil nicht dadurch zunichte macht, dass man die Daten ungeschickt anordnet und sie dann stets umspeichern muß.

Für die Übungen wird eine zweidimensionale FFT-Routine aus [2] bereitgestellt. Die Ein- und Ausgabedaten sind als Arrays von komplexen Zahlen angeordnet; jede komplexe Zahl als *struct* aus zwei *double* (nicht wie manchmal als *float*). Der Definitionsbereich von \hat{f} ist $[0, N_x - 1] \times [0, N_y - 1]$.

7.2 Faltungsoperatoren

7.2.1 Die Faltung auf \mathbb{R}^n

Für $f, g \in L_1(\mathbb{R}^n)$ existiert $f \star g(p) := \int_{\mathbb{R}^n} f(p-q)g(q) dq$ für alle $p \in \mathbb{R}^n$ außer auf einer Lebesgue-Nullmenge, und es gilt $f \star g \in L_1(\mathbb{R}^n)$.

Der Operator $\star : L_1(\mathbb{R}^n) \times L_1(\mathbb{R}^n) \to L_1(\mathbb{R}^n)$ heißt die Faltung auf \mathbb{R}^n . $f \star g$ heißt Faltung von f mit g.

7.2.1.1 Eigenschaften. Seien $f, g, h \in L_1(\mathbb{R}^n)$.

- a) Die Faltung ist bilinear und stetig.
- b) Die Faltung ist kommutativ, das heißt, es gilt $f \star g = g \star f$.
- c) Die Faltung ist assoziativ, das heißt, es gilt $(f \star g) \star h = f \star (g \star h)$.
- d) Mit der Faltung als Multiplikation ist $L_1(\mathbb{R}^n)$ eine kommutative Banachalgebra.
- e) Ist f stetig differenzierbar, so gilt $\frac{\partial}{\partial x_j}(f \star g) = \frac{\partial f}{\partial x_j} \star g$, für jedes j.
7.2.2 Die Faltung auf \mathbb{Z}^n

 $L_1(\mathbb{Z}^n)$ sei der Vektorraum der Abbildungen $f:\mathbb{Z}^n\to\mathbb{C}$ mit $\sum_{p\in\mathbb{Z}^n}|f(p)|:=\|f\|_1<\infty$.

Für $f, g \in L_1(\mathbb{Z}^n)$ existiert $f \star g(p) := \sum_{q \in \mathbb{Z}^n} f(p-q)g(q)$ für jedes $p \in \mathbb{Z}^n$, und es gilt $f \star g \in L_1(\mathbb{Z}^n)$. Der Operator $\star : L_1(\mathbb{Z}^n) \times L_1(\mathbb{Z}^n) \to L_1(\mathbb{Z}^n)$ heißt die *Faltung* auf \mathbb{Z}^n .

Die Eigenschaften 7.2.1.1 a)-d) gelten entsprechend.

7.2.3 Die zyklische Faltung

Sei $G = \mathbb{Z}_{N_1} \times \ldots \times \mathbb{Z}_{N_n}$. Weil G endlich ist, ist jede Funktion auf G absolut summierbar und somit ist $L_1(G)$ einfach der Raum aller Funktionen $f : G \to \mathbb{C}$. Daher ist die Faltung auf G die Abbildung

$$\star: \mathbb{C}^G \times \mathbb{C}^G \to \mathbb{C}^G, \ (f,g) \mapsto f \star g \text{ mit } f \star g(p) = \sum_{q \in G} f(p-q)g(q).$$

Die Eigenschaften 7.2.1.1 a)-d) gelten entsprechend.

Für praktische Rechnungen wählt man für G ein Repräsentantensystem. Wir betrachten zunächst den eindimensionalen Fall, also $G = \mathbb{Z}/N\mathbb{Z}$, und wählen als Repräsentantensystem ein Intervall $R \subset \mathbb{Z}$ der Länge N.

Für $f: R \to \mathbb{C}$ und $g: R \to \mathbb{C}$ ist dann $f \star g: R \to \mathbb{C}$ definiert als

$$f \star g(t) = \sum_{s=0}^{N-1} f(t-s)g(s)$$

wobei die Argumente modulo N so zu wählen sind, daß sie in R liegen.

Ist also z.B. R = [0, N - 1], so werden die Punkte $N, N + 1, \ldots$ mit $0, 1, \ldots$ identifiziert, ebenso $-1, -2, \ldots$ mit $N - 1, N - 2, \ldots$ Deshalb spricht man von zyklischer Faltung.

Die Situation im *n*-dimensionalen Fall, also $G = \mathbb{Z}_{N_1} \times \ldots \times \mathbb{Z}_{N_n}$, ist völlig analog; man rechnet eben in der *j*-ten Komponente der Argumente stets modulo N_j und wählt so immer einen Punkt im Repräsentantensystem R.

7.2.4 Definition von Faltungsoperatoren

7.2.4.1 Definition. Set $G = \mathbb{R}^n$ oder $G = \mathbb{Z}^n$ oder $G = \mathbb{Z}_{N_1} \times \ldots \times \mathbb{Z}_{N_n}$.

Eine Abbildung $T: L_1(G) \to L_1(G)$ heißt Faltungsoperator, wenn es ein $h \in L_1(G)$ gibt, so daß $Tf = h \star f$ für jedes $f \in L_1(G)$. Einen Faltungsoperator auf $\mathbb{Z}_{N_1} \times \ldots \times \mathbb{Z}_{N_n}$ nennt man auch zyklischen Faltungsoperator.

h heißt die *Impulsantwort* von *T*. Der Name erklärt sich aus folgender Eigenschaft: Ist $G = \mathbb{Z}^n$ oder $G = \mathbb{Z}_{N_1} \times \ldots \times \mathbb{Z}_{N_n}$, so gilt $h = T\delta_0$, wobei $\delta_0(p) = 1$ für p = 0, und 0 sonst, der *Einheitsimpuls* im Ursprung ist.

7.2.4.2 Bemerkung. Im Falle $G = \mathbb{R}^n$ ist die Funktion f(p) = 1 für p = 0, und 0 sonst, fast überall 0 und repräsentiert somit das Nullelement in $L_1(\mathbb{R}^n)$. Man kann aber die Faltung ausdehnen auf den Raum der endlichen Maße und dann gilt $h = h \star \delta_0 = T\delta_0$, wobei δ_0 das Diracmaß im Ursprung ist.

7.2.4.3 Definition. Ein Faltungsoperator auf \mathbb{R}^n oder \mathbb{Z}^n heißt ein *FIR-Filter (finite impulse response)*, wenn seine Impulsantwort beschränkten Träger hat, das heißt, außerhalb einer beschränkten Menge verschwindet.

7.2.4.4 Bemerkung. FIR-Filter lassen sich problemlos auf einen größeren Definitionsbereich fortsetzen, der alle beschränkten Signale umfaßt.

- $G = \mathbb{Z}^n$: Hat $h : \mathbb{Z}^n \to \mathbb{C}$ beschränkten Träger, so ist $h \star f(p) = \sum_{q \in \mathbb{Z}^n} h(p-q)f(q)$ für jede Funktion $f : \mathbb{Z}^n \to \mathbb{C}$ und jedes $p \in \mathbb{Z}^n$ wohldefiniert, weil nur endlich viele Summanden nicht verschwinden.
- $G = \mathbb{R}^n$: Hat $h \in L_1(\mathbb{R}^n)$ beschränkten Träger, so ist $h \star f(p) = \int_{\mathbb{R}^n} h(p-q)f(q) dq$ für jedes $f \in L_\infty(\mathbb{R}^n)$ und fast jedes $p \in \mathbb{R}^n$ wohldefiniert.

Fazit: FIR-Filter sind auf $L_{\infty}(\mathbb{R}^n)$ definiert.

7.2.5 Einige Eigenschaften von Faltungsoperatoren

- a) Faltungsoperatoren sind linear.
- b) Faltungsoperatoren sind translationsinvariant, das heißt, sie kommutieren mit Translationen.
- c) Faltungsoperatoren kommutieren miteinander, das heißt, es gilt $T_1 \circ T_2 = T_2 \circ T_1$.

7.2.6 Lineare Filter und Faltungsoperatoren auf \mathbb{Z}^2

Wir verwenden die Bezeichnungen aus 2.4. Sei also $M = \llbracket -n_x, n_x \rrbracket \times \llbracket -n_y, n_y \rrbracket$ ein Operatorfenster und $\varphi : \mathbb{C}^M \to \mathbb{C}$ eine lineare, lokale Filteroperation mit Matrix $A = (a(k, l))_{(k,l) \in M}$. Dann gilt für das davon erzeugte lineare Filter T_{φ}

$$T_{\varphi}f(x,y) = \sum_{(k,l)\in M} a(k,l)f(x+k,y+l) \text{ für alle } (x,y) \text{ mit } (x,y) + M \subset R.$$

Wir betrachten hier nur Eingangssignale f, deren Definitionsbereich R gleich ganz \mathbb{Z}^2 ist.

 $\mathbb{B}(\mathbb{Z}^2,\mathbb{C})$ sei der Vektorraum aller Funktionen $f:\mathbb{Z}^2\to\mathbb{C}$.

7.2.6.1 Lemma. Ein Operator $\mathbb{B}(\mathbb{Z}^2, \mathbb{C}) \to \mathbb{B}(\mathbb{Z}^2, \mathbb{C})$ wird genau dann von einer linearen, lokalen Filteroperation erzeugt, wenn er ein Faltungsoperator ist, dessen Impulsantwort beschränkten Träger hat.

Beweis.

⇒: Sei φ eine lineare Filteroperation mit Matrix $A = (a(k, l))_{(k,l) \in M} \in \mathbb{C}^M$. Setze $\tilde{a} : \mathbb{Z}^2 \to \mathbb{C}$, $\tilde{a}(k, l) := a(-k, -l)$ für $(k, l) \in M$ und $\tilde{a} := 0$ sonst. Dann gilt:

$$T_{\varphi}f(x,y) = \sum_{(k,l) \in M} a(k,l)f(x+k,y+l) = \sum_{(k,l) \in \mathbb{Z}^2} \tilde{a}(-k,-l)f(x+k,y+l) = f \star \tilde{a}(x,y) = \tilde{a} \star f(x,y).$$

 $\begin{array}{ll} \Leftarrow: & \text{Sei } T \text{ ein Faltungsoperator mit Impulsantwort } h \text{ mit beschränktem Träger. Dann gibt es ein Rechteck} \\ & M = \llbracket -n_x, n_x \rrbracket \times \llbracket -n_y, n_y \rrbracket, \text{ außerhalb dessen } h \text{ verschwindet.} \\ & \text{Setze } a(k,l) = h(-k,-l) \text{ für jedes } (k,l) \in M \text{ und } A = (a(k,l))_{(k,l) \in M} \in \mathbb{C}^M. \\ & \text{Dann gilt: } T_A f = h \star f \text{ für jedes } f \in \mathbb{B}(\mathbb{Z}^2,\mathbb{C}). \end{array}$

7.2.7 Lineare Filter und zyklische Faltung

In Anwendungen sind die Bildsignale auf einem beschränkten Rechteck $R = [0, N_x - 1] \times [0, N_y - 1]$ definiert, und lineare, lokale Filteroperationen sind in der Nähe des Randes von R nicht oder nur etwas willkürlich definiert. Eine einfache Möglichkeit, eine lineare, lokale Filteroperation in Randnähe zu definieren, besteht darin, in den Argumenten modulo N_x bzw. N_y zu rechnen, so daß man stets einen Punkt in R erhält.

Wir sprechen dann von einer zyklischen, linearen, lokalen Filteroperation.

Dabei sei stets stillschweigend $n_x < \frac{N_x}{2}$ und $n_y < \frac{N_y}{2}$ vorausgesetzt.

7.2.7.1 Lemma. Ein Operator $\mathbb{C}^R \to \mathbb{C}^R$ wird genau dann von einer zyklischen, linearen, lokalen Filteroperation erzeugt, wenn er durch zyklische Faltung mit einer Impulsantwort $h: R \to \mathbb{C}$ entsteht.

Beweis.

Wie zu 7.2.6.1, nur wird in den Argumenten modulo N_x bzw. N_y gerechnet.

Verdeutlichung, wie die Impulsantwort $h: R \to \mathbb{C}$ aus der Filtermaske $a: M \to \mathbb{C}$ entsteht:

- 1) Gehe zu $a_-: M \to \mathbb{C}$ über mit $a_-(k, l) = a(-k, -l)$.
- 2) h wird in den Quadranten wie a_{-} definiert und sonst 0 gesetzt.



Oder: Biege R zu einem Torus zusammen durch Identifizieren gegenüberliegender Seiten.

7.2.7.2 Bemerkung. Die zyklische Definition linearer, lokaler Filteroperationen bietet zwar rechentechnisch Vorteile; sie liefert aber in Randnähe meist unsinnige Resultate, weil die Bildstrukturen an gegenüberliegenden Rändern meist nichts miteinander zu tun haben.

7.3 Beschreibung von Faltungsoperatoren im Frequenzbereich

7.3.1 Satz. Sei $R = [\![0, N_x - 1]\!] \times [\![0, N_y - 1]\!]$.

- a) Für $f, g \in L_1(\mathbb{R}^n)$ gilt $\widehat{f \star g} = \widehat{f} \cdot \widehat{g}$ und $\widehat{f \cdot g} = \widehat{f} \star \widehat{g}$, falls $f \cdot g, \widehat{f}, \widehat{g}, \widehat{f \cdot g} \in L_1(\mathbb{R}^n)$.
- b) Für $f, g \in \mathbb{C}^R$ gilt $\widehat{f \star g} = \hat{f} \cdot \hat{g}$ und $\widehat{f \cdot g} = \frac{1}{N_\pi N_\nu} \hat{f} \star \hat{g}$.

7.3.2 Korollar.

- a) Ist T ein Faltungsoperator auf \mathbb{R}^n mit Impulsantwort $h \in L_1(\mathbb{R}^n)$, so gilt $Tf = \mathcal{F}^{-1}(\hat{h} \cdot \hat{f})$.
- b) Ist T ein zyklischer Faltungsoperator auf R mit Impulsantwort $h: R \to \mathbb{C}$, so gilt $Tf = \mathcal{F}^{-1}(\hat{h} \cdot \hat{f})$.

In beiden Fällen wird \hat{h} die Übertragungsfunktion oder Systemfunktion von T genannt. Oft wird die Impulsantwort mit einem kleinen Buchstaben und die Übertragungsfunktion mit dem entsprechenden Großbuchstaben bezeichnet.

7.3.3 Bemerkung. Aus der rechenintensiven Faltungsoperation wird im Frequenzbereich also einfach die punktweise Multiplikation mit der Übertragungsfunktion. Weil es für die DFT und die inverse DFT schnelle Algorithmen gibt, ist die Berechnung von Faltungsoperatoren über dem Frequenzbereich oft schneller als im Ortsbereich.

7.3.4 Eine abstrakte Deutung

Im diskreten Fall wird T bezüglich der Basis $(\delta_{(u,v)})_{(u,v)\in R}$ (als linearer Operator) durch eine $N_x N_y \times N_x N_y$ -Matrix beschrieben (die aus der Impulsantwort h berechnet werden kann). Bezüglich der Basis $(E_{(u,v)})_{(u,v)\in R}$ hat T eine einfachere Darstellung, denn im diskreten Fall gilt $f = \frac{1}{N_x N_y} \sum_{(k_x,k_y)} \hat{f}(k_x,k_y) E_{(k_x,k_y)}$ und

$$Tf = \frac{1}{N_x N_y} \sum_{(k_x, k_y)} \widehat{Tf}(k_x, k_y) E_{(k_x, k_y)} = \frac{1}{N_x N_y} \sum_{(k_x, k_y)} \hat{h}(k_x, k_y) \hat{f}(k_x, k_y) E_{(k_x, k_y)} \tag{*}$$

Nach 7.1.3.5.f) gilt $\widehat{E_{(u,v)}} = N_x N_y \delta_{(u,v)}$ und somit nach (*)

$$TE_{(u,v)} = \hat{h}(u,v)E_{(u,v)}$$
 (**)

Dies bedeutet, daß jedes $E_{(u,v)}$ Eigenvektor von T zum Eigenwert $\hat{h}(u,v)$ ist. Die $(E_{(u,v)})_{(u,v)\in R}$ sind also für *jeden* Faltungsoperator $\mathbb{C}^R \to \mathbb{C}^R$ eine Basis aus Eigenvektoren. Und bezüglich einer Basis aus Eigenvektoren wird ein linearer Operator durch eine Diagonalmatrix $\in \mathbb{C}^{R \times R}$ beschrieben. Die Diagonalelemente sind gerade die Eigenwerte, die nach (**) die Werte der Übertragungsfunktion sind.

7.3.5 Eine anschauliche Deutung

Ein Faltungsoperator mit Impulsantwort h zerlegt das Eingangssignal f mit der Fouriertransformation in harmonische Schwingungen $\hat{f}(\omega_x, \omega_y) e^{2\pi i (\omega_x x + \omega_y y)}$ bzw. $\hat{f}(k_x, k_y) E_{(k_x, k_y)}$, verändert Amplitude und Phase durch Multiplikation mit $\hat{h}(\omega_x, \omega_y)$ bzw. $\hat{h}(k_x, k_y)$ und setzt aus diesen Schwingungen mit der inversen Fouriertransformation das Ausgangssignal zusammen.

Experimentell stellt man fest, daß die Bildinformation wesentlich durch das Phasenspektrum bestimmt wird. Veränderungen des Amplitudenspektrums wirken sich ähnlich wie die Filteroperationen aus, die wir schon im 3. Kapitel betrachtet haben. Filter, die nur das Phasenspektrum, aber nicht das Amplitudenspektrum verändern, sind zwar für manche Anwendung wichtig, werden aber hier nicht näher untersucht.

Faltungsoperatoren (oder lineare Filter) werden oft nach der Art klassifiziert, wie sie das Amplitudenspektrum verändern.

7.3.6 Filtercharakteristiken

T sei ein Faltungsoperator mit Impulsantwort h und Übertragungsfunktion \hat{h} .

T heißt Tiefpaßfilter, wenn $|\dot{h}|$ in einer Umgebung des Ursprungs deutlich größere Werte annimmt als außerhalb; wenn also T tieferfrequente Schwingungen passieren läßt, höhere aber dämpft.



Typischer Verlauf von $|\hat{h}|$ eines Tiefpaß-Filters im eindimensionalen, kontinuierlichen Fall

T heißt Hochpaßfilter, wenn $|\hat{h}|$ in einer Umgebung des Ursprungs deutlich kleinere Werte annimmt als außerhalb; wenn also T höherfrequente Schwingungen passieren läßt, tiefere aber dämpft.



Typischer Verlauf von $|\hat{h}|$ eines Hochpaß-Filters im eindimensionalen, kontinuierlichen Fall

Theißt $Bandpa\betafilter$, wenn Thohe und tiefe Frequenzen dämpft und nur die in einem kleinen Bereich passieren läßt.

Typischer Verlauf von $|\hat{h}|$ eines Bandpaß-Filters im eindimensionalen, kontinuierlichen Fall

T heißt Sperrfilter, wenn T alle Frequenzbereiche passieren läßt außer denen in einem kleinen Bereich.

Typischer Verlauf von $|\hat{h}|$ eines Sperr-Filters im eindimensionalen, kontinuierlichen Fall

Der Verlauf von $|\hat{h}|$ muß nicht symmetrisch zum Ursprung sein. Soll aber Tf stets reell sein, wenn f reell ist, so muß \hat{h} konjugiert komplex symmetrisch sein und somit $|\hat{h}|$ symmetrisch.

Obige Filterklassifikation ist im Diskreten nur dann sinnvoll, wenn man als Definitionsbereich von \hat{h} das Intervall $\left[\left[-\frac{N_x}{2}, \frac{N_x}{2} - 1\right]\right]$ bzw. das Rechteck $\left[\left[-\frac{N_x}{2}, \frac{N_x}{2} - 1\right]\right] \times \left[\left[-\frac{N_y}{2}, \frac{N_y}{2} - 1\right]\right]$ wählt.

7.3.7 Beispiele kontinuierlicher Filter in einer Dimension

1) Das ideale Tiefpaßfilter:

$$\hat{h}(\omega) := \operatorname{rect}_{2\omega_0}(\omega) := \begin{cases} 1 & \text{, für } |\omega| < |\omega_0| \\ 0 & \text{, sonst} \end{cases}$$

Es gilt $h(t) = \mathcal{F}^{-1}(\hat{h})(t) = \frac{\sin(2\omega_0\pi t)}{\pi t}$ (vgl. Beispiel 1 aus 7.1.1). Es ist kein FIR-Filter. Analog definiert man ideale Hoch- und Bandpaß-Filter. Auch sie sind keine FIR-Filter.

2) Das Gaußsche Tiefpaßfilter:

$$\hat{h}(\omega) := \exp\left(-\frac{\omega^2}{2\alpha^2}\right).$$

Es gilt $h(t) = \sqrt{2\pi}\alpha \exp(-2\pi^2\alpha^2 t^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{t^2}{2\sigma^2})$. Dies ist eine (normierte) Gaußfunktion mit der Streuung $\sigma = \frac{1}{2\pi\alpha}$. Auch dies ist kein FIR-Filter. Weil aber die Gaußfunktion schnell abfällt, kann man sie z.B. außerhalb $[-3\sigma, 3\sigma]$ ab-

weil aber die Gaubrunktion schneil abfallt, kann man sie z.B. aubernalb $[-3\sigma, 3\sigma]$ abschneiden, das heißt, auf Null setzen, und erhält somit eine gute Approximation durch ein FIR-Filter.

3) Das Boxfilter:

$$h(t) := \operatorname{rect}_a(t) := \begin{cases} 1 & , \text{ für } |t| < \frac{a}{2} \\ 0 & , \text{ sonst} \end{cases}.$$

Es gilt $\hat{h}(\omega) = a \frac{\sin a\pi\omega}{a\pi\omega}$. Das Boxfilter hat also Tiefpaßcharakteristik. Aber die Dämpfung nimmt nicht monoton mit wachsender Frequenz zu. Außerdem nimmt die Höhe der lokalen Maxima nur langsam ab.

4) Differentiation:

Nach Beispiel 7.1.1.3) gilt
$$\frac{\widehat{df}}{dt}(\omega) = 2\pi i \omega \widehat{f}(\omega)$$
, also $\frac{\widehat{d^n f}}{dt^n}(\omega) = (2\pi i)^n \omega^n \widehat{f}(\omega)$.

Ein linearer Differentialoperator mit konstanten Koeffizienten entspricht im Frequenzbereich also der Multiplikation mit einem Polynom und hat somit Hochpaßcharakteristik. (Er ist kein Faltungsoperator im Sinne von 7.2.4, wohl aber in einem allgemeineren Sinn.)

5) Komposition von Gauß- und Ableitungsfiltern: Erste Ableitung:

ł

$$a(t) := \frac{d}{dt} \left(\frac{1}{\sqrt{2\pi\sigma}} g_{\sigma}(t) \right), \ g_{\sigma}(t) := \exp\left(-\frac{t^2}{2\sigma^2} \right).$$

Es gilt $\hat{h}(\omega) = \frac{2\pi i}{\sqrt{2\pi\sigma}} \omega \hat{g}_{\sigma}(\omega) = 2\pi i \omega g_{\alpha}(\omega)$ mit $\alpha = \frac{1}{2\pi\sigma}$.

Beachte: $\frac{d}{dt}(g_{\sigma} \star f) = \frac{dg_{\sigma}}{dt} \star f$ nach 7.2.1.1.e). Das Filter hat Bandpaßcharakteristik. Zweite Ableitung:

$$\frac{d^2}{dt^2}(g_{\sigma}\star f) = \frac{d^2g_{\sigma}}{dt^2}\star f, \ h(t) = g_{\sigma}'', \ \hat{h}(\omega) = -4\pi^2\omega^2\hat{g}_{\alpha}(\omega) = -4\pi^2\omega^2g_{\alpha}(\omega) \text{ mit } \alpha = \frac{1}{2\pi\sigma}.$$

Der Filter besitzt Bandpaßcharakteristik.

6) Gaborfilter:

$$h(\omega) := g_{\alpha}(\omega - \lambda)$$

Es gilt $h(t) = \frac{1}{\sqrt{2\pi\sigma}}g_{\sigma}(t)e^{2\pi i\lambda t}$ mit $\sigma = \frac{1}{2\pi\alpha}$, dabei spricht man von der *Mittenfrequenz* λ und der *Bandbreite* α .

Das Filter hat Bandpaßcharakteristik. Seine Impulsantwort ist komplexwertig.

Anschauliche Deutung: Betrag und Winkel von $h \star f(t)$ geben an, wie stark f in der Nähe von t einer Schwingung mit Frequenz circa λ ähnelt und mit welcher Phasenlage. Re(h) ist eine exponentiell abklingende Cosinusschwingung, Im(h) eine exponentiell abklingende Sinunsschwingung.











7.3.8 Beispiele kontinuierlicher Filter in zwei Dimensionen

Die Analyse vieler zweidimensionaler Filter läßt sich auf die eindimensionaler Filter zurückführen, weil die Filter separabel sind: Ist $h(x, y) = h_x(x) \cdot h_y(y)$ mit $h_x, h_y \in L_1(\mathbb{R})$, so gilt $\mathcal{F}_2h(\omega_x, \omega_y) = \mathcal{F}_1h_x(\omega_x) \cdot \mathcal{F}_1h_y(\omega_y)$, wobei \mathcal{F}_n die *n*-dimensionale Fouriertransformation ist.

- 1) Ideale Tiefpaßfilter: $\hat{h} = \chi_{\mathcal{U}}$, wobei \mathcal{U} eine Nullumgebung ist.
 - a) $\mathcal{U} = [-a_x, a_x] \times [-a_y, a_y]$ Rechteck $(a_x, a_y > 0)$; also $\hat{h}(\omega_x, \omega_y) = \operatorname{rect}_{2a_x}(\omega_x) \cdot \operatorname{rect}_{2a_y}(\omega_y)$. Dann folgt

$$h(x,y) = \frac{\sin 2a_x \pi x}{\pi x} \frac{\sin 2a_y \pi y}{\pi y}.$$

Das Filter ist kein FIR-Filter.

- b) $\mathcal{U} = B(0, a)$ Kreisscheibe (a > 0). Dann ist das Filter nicht separabel. Man kann zeigen, daß $h(x, y) = \varphi(\sqrt{x^2 + y^2})$, wobei $\varphi : [0, \infty[\rightarrow \mathbb{R}$ eine Funktion ist, die ähnlich oszilliert wie $\frac{\sin at}{at}$ (allerdings sind die Nullstellen nicht exakt äquidistant; Stichwörter: Hankeltransformation, Besselfunktion 1. Art).
- 2) Das Gaußsche Tiefpaßfilter: $\hat{h}(\omega_x, \omega_y) = e^{-\frac{\omega_x^2}{2\alpha_x^2}} e^{-\frac{\omega_y^2}{2\alpha_y^2}}$. Im isotropen Fall gilt $\alpha_x = \alpha_y = \alpha$, also $\hat{h}(\omega_x, \omega_y) = e^{-\frac{1}{2\alpha^2}(\omega_x^2 + \omega_y^2)} = g_\alpha(\sqrt{\omega_x^2 + \omega_y^2})$.

Das Filter ist separabel; somit gilt

$$h(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{y^2}{2\sigma_y^2}}$$

mit $\sigma_x = \frac{1}{2\pi\alpha_x}$ und $\sigma_y = \frac{1}{2\pi\alpha_y}$; und im isotropen Fall mit $\sigma = \frac{1}{2\pi\alpha}$

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x^2 + y^2)} =: g_{2,\sigma}(x,y).$$

3) Gaborfilter (isotrop): $\hat{h}(\omega_x, \omega_y) = e^{-\frac{1}{2\alpha^2} \left((\omega_x - \lambda_x)^2 + (\omega_y - \lambda_y)^2 \right)} = g_\alpha(\|\omega - \lambda\|)$. Das Filter ist separabel, und es gilt

$$h(x,y) = \frac{1}{2\pi\sigma^2} g_{\sigma}(\sqrt{x^2 + y^2}) e^{2\pi i (\lambda_x x + \lambda_y y)} = \frac{1}{2\pi\sigma^2} g_{\sigma}(\|(x,y)\|) e^{2\pi i \langle (\lambda_x,\lambda_y), (x,y) \rangle} \text{ mit } \sigma = \frac{1}{2\pi\sigma^2} g_{\sigma}(\|(x,y)\|) e^{2\pi i \langle (\lambda_x,\lambda_y), (x,y) \rangle}$$

4) LOG- und DOG-Filter: Sei $g_{2,\sigma}(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$. LOG = Laplacian of Gaussian: $h_{LOG} = \Delta g_{2,\sigma} = \frac{\partial^2}{\partial x^2}g_{2,\sigma} + \frac{\partial^2}{\partial y^2}g_{2,\sigma}$. DOG = Difference of Gaussians: $h_{DOG} = g_{2,\sigma_1} - g_{2,\sigma_2}$ mit $\sigma_1 > \sigma_2$. h_{LOG} und h_{DOG} sind rotationsinvariant. Es gilt:

$$\hat{h}_{LOG}(\omega_x, \omega_y) = -4\pi^2 (\omega_x^2 + \omega_y^2) \cdot \hat{g}_{2,\sigma}(\omega_x, \omega_y) = -4\pi^2 (\omega_x^2 + \omega_y^2) \cdot g_\alpha(\sqrt{\omega_x^2 + \omega_y^2})$$

$$= -4\pi^2 \|(\omega_x, \omega_y)\|^2 \cdot g_\alpha(\|(\omega_x, \omega_y)\|) \text{ mit } \alpha = \frac{1}{2\pi\sigma}.$$

Auf jedem Strahl durch den Ursprung sieht die Übertragungsfunktion aus wie die von g''_{σ} im eindimensionalem (siehe 7.3.7). Weiter gilt:

$$\hat{h}_{DOG}(\omega_x, \omega_y) = \hat{g}_{2,\sigma_1}(\omega_x, \omega_y) - \hat{g}_{2,\sigma_2}(\omega_x, \omega_y) = g_{\alpha_1}(\|(\omega_x, \omega_y)\|) - g_{\alpha_2}(\|(\omega_x, \omega_y)\|)$$

$$\stackrel{(if \alpha_1, \dots, \alpha_n)}{=} \frac{1}{2} \sum_{\alpha_1, \alpha_2, \dots, \alpha_n} \frac{1}{2} \sum_{\alpha_1, \dots, \alpha_n} \frac{1}{2} \sum_{\alpha_1, \dots,$$

mit $\alpha_1 = \frac{1}{2\pi\sigma_1}$, $\alpha_2 = \frac{1}{2\pi\sigma_2}$, $g_\alpha(t) = e^{-\frac{t^2}{2\alpha^2}}$ (nicht normiert).

Auf jedem Strahl durch den Ursprung ist die Übertragungsfunktion die Differenz zweier eindimensionaler Gaußglocken (nicht normiert). Auch \hat{h}_{LOG} und \hat{h}_{DOG} sind rotationsinvariant.

LOG- und DOG-Filter haben Bandpaßcharakteristik. Ihre Impulsantworten und Übertragungsfunktionen sind qualitativ sehr ähnlich. Bei geeigneter Wahl von σ_1 und σ_2 stimmt das DOG-Filter sogar quantitativ sehr gut mit einem LOG-Filter überein ($\sigma_1 \approx 1.6\sigma_2$ und $\sigma_2 \approx 0.8\sigma$).

7.3.9 Beispiele diskreter Filter in zwei Dimensionen

1) 3×3 -Glättungsfilter mit Kern $A = (a_1, a_0, a_1)^T \cdot (b_1, b_0, b_1)$. Sei $M = \llbracket -1, 1 \rrbracket^2$ und $A = (a(x, y)) \in \mathbb{R}^M$. h(x, y) = a(-x, -y) für $(x, y) \in M$ und 0 sonst sei die Impulsantwort. Es gilt:

$$\hat{h}(k_x, k_y) = \sum_x \sum_y h(x, y) e^{-\frac{2\pi i}{N_x} k_x x} e^{-\frac{2\pi i}{N_y} k_y y}$$

= $\left(a_0 + a_1 e^{-\frac{2\pi i}{N_y} k_y} + a_1 e^{\frac{2\pi i}{N_y} k_y}\right) \cdot \left(b_0 + b_1 e^{-\frac{2\pi i}{N_x} k_x} + b_1 e^{\frac{2\pi i}{N_x} k_x}\right)$
= $\left(a_0 + 2a_1 \cos \frac{2\pi}{N_y} k_y\right) \cdot \left(b_0 + 2b_1 \cos \frac{2\pi}{N_x} k_x\right)$

Rechteckfilter: $a_0 = a_1 = b_0 = b_1 = \frac{1}{3}$. Es gilt:

$$\hat{h}(k_x, k_y) = \frac{1}{9} \left(1 + 2\cos\frac{2\pi}{N_x} k_x \right) \cdot \left(1 + 2\cos\frac{2\pi}{N_y} k_y \right)$$

Das Rechteckfilter hat also Tiefpaßcharakteristik, allerdings keine sehr gute, da die Dämpfung nicht monoton mit der Frequenz ansteigt. (Vergleiche das in 2.4.1.1 beschriebene Phänomen und die Charakteristik des kontinuierlichen Rechteckfilters in 7.3.7.)

Binomialfilter: $A = \frac{1}{16}(1, 2, 1)^T \cdot (1, 2, 1)$. Es gilt:



2) 3×3 -Differenzenfilter mit Kern $A = (1, 1, 1)^T \cdot (-1, 0, 1)$.

$$\hat{h}(k_x, k_y) = \left(1 + 2\cos\frac{2\pi}{N_y}k_y\right) \cdot 2i\sin\frac{2\pi}{N_x}k_x.$$

7.4 Bemerkungen zur Konstruktion diskreter Filter

7.4.1 Entwurf im Frequenzbereich

Man gibt sich einfach eine Übertragungsfunktion $H_d : [\![-\frac{N_x}{2}, \frac{N_x}{2} - 1]\!] \times [\![-\frac{N_y}{2}, \frac{N_y}{2} - 1]\!] \to \mathbb{C}$ mit der gewünschten Filtercharakteristik vor. Dies kann z.B. durch Abtastung einer Übertragungsfunktion $H_c : \mathbb{R}^2 \to \mathbb{C}$ im Kontinuierlichen geschehen auf folgende Weise:

Ein kontinuierliches Filter mit Übertragungsfunktion H_c verändert ein Signal f_c dadurch, daß jeder Anteil

$$f_c(\omega_x, \omega_y) e^{2\pi i \omega_x x} e^{2\pi i \omega_y y}$$
 mit $H_c(\omega_x, \omega_y)$

multipliziert wird.

Ein diskretes Filter mit Übertragungsfunktion H_d verändert ein Signal f_d dadurch, daß jeder Anteil

$$\hat{f}_d(k_x, k_y) e^{\frac{2\pi i}{N_x}k_x x} e^{\frac{2\pi i}{N_y}k_y y}$$
 mit $H_d(k_x, k_y)$

multipliziert wird.

Die diskrete Schwingung

$$E_{(k_x,k_y)}: R \to \mathbb{C}, \ E_{(k_x,k_y)}(x,y) = e^{\frac{2\pi i}{N_x}k_x x} e^{\frac{2\pi i}{N_y}k_y y}$$

ist die Einschränkung der kontinuierlichen Schwingung

$$\mathbb{R}^2 \to \mathbb{C}, \ (x,y) \mapsto e^{2\pi i \omega_x x} e^{2\pi i \omega_y y}$$

mit $\omega_x = \frac{k_x}{N_x}$ und $\omega_y = \frac{k_y}{N_y}$.

Setzt man $H_d(k_x, k_y) := H_c(\frac{k_x}{N_x}, \frac{k_y}{N_y})$, so werden die Signalanteile mit Frequenz (k_x, k_y) bzw. $(\frac{k_x}{N_x}, \frac{k_y}{N_y})$ von den diskreten bzw. dem kontinuierlichen Filter in gleicher Weise verändert. **Bemerkung** Dies läßt sich mathematisch schöner und präziser im Bahmen von Distributionen formulieren:

Bemerkung. Dies läßt sich mathematisch schöner und präziser im Rahmen von Distributionen formulieren; vergleiche dazu die Vorlesung über Signalverarbeitung.

7.4.1.1 Beispiel.

- 1) Diskrete Gauß-, Gabor- und LOG-Filter kann man durch entsprechende Abtastung der Übertragungsfunktion der kontinuierlichen Filter erhalten.
- 2) Man kann die Differentiation in gewissem Sinn exakt ins Diskrete übertragen: Im Kontinuierlichen entspricht der Ableitung $\frac{\partial}{\partial x}$ im Frequenzbereich die Multiplikation mit $2\pi i\omega_x$. Setzt man $H_d(k_x, k_y) := 2\pi i \frac{k_x}{N_x}$, so bewirkt das diskrete Filter mit Übertragungsfunktion H_d tatsächlich, daß jede Schwingung $E_{(k_x,k_y)}$ zu $2\pi i \frac{k_x}{N_x} E_{(k_x,k_y)} = \frac{\partial}{\partial x} E_{(k_x,k_y)}$ verändert wird.

7.4.1.2 Probleme.

- 1) Definiert man ein diskretes Filter im Frequenzbereich durch Multiplikation mit einer Übertragungsfunktion H, so entspricht dem im Ortsbereich die zyklische Faltung mit $h = \mathcal{F}^{-1}(H)$. Man beachte, daß h i.a. keinen kleinen Träger haben wird, daß man also i.a. keine lokale Filteroperation erhält und somit eine Realisierung im Ortsbereich ineffizient sein kann (außer wenn h sehr schnell abklingt und abgeschnitten werden kann, ohne einen größeren Fehler zu machen).
- 2) Bei Vorgabe einer beliebigen Übertragungsfunktion H ist nicht gewährleistet, daß das Filter aus einem Bild f wieder ein Bild macht, das heißt, daß $\mathcal{F}^{-1}(H \cdot \hat{f})$ wieder reell und nicht negativ ist. Daß das Ergebnis wieder reell ist, kann man erzwingen, indem man H konjugiert komplex symmetrisch wählt (dies ist auch notwendig).

7.4.2 Diskretisierung kontinuierlicher Filter im Ortsbereich

Die Eigenschaft kontinuierlicher Filter sind oft einfacher zu analysieren als die diskreter, weil mächtige Hilfsmittel der Infinitesimalrechnung zur Verfügung stehen. Deshalb geht man bei der Konstruktion diskreter Filter oft so vor, daß man zunächst ein geeignetes kontinuierliches Filter konstruiert und dessen Impulsantwort diskretisiert.

 $h: \mathbb{R}^2 \to \mathbb{C}$ sei die Impulsantwort eines kontinuierlichen Filters T und $M = \llbracket -n_x, n_x \rrbracket \times \llbracket -n_y, n_y \rrbracket$. Weiter sei h integrierbar und in den Punkten von M stetig. Für stetiges $f \in L_1(\mathbb{R}^2)$ ist

$$\sum_{(u,v)\in M} f(u,v) h(x-u,y-v) \quad \text{eine Riemannsche Summe zu}$$
$$h \star f(x,y) = \int_{\mathbb{R}^2} f(u,v) h(x-u,y-v) \, du dv.$$

Wir setzen jetzt voraus, daß

- h innerhalb eines Quadrates der Seitenlänge 1 nicht sehr schwankt
- h außerhalb von $M' := [-n_x, n_x] \times [-n_y, n_y]$ sehr kleine Werte annimmt, so daß $\int_{\mathbb{R}^2 \setminus M'} |h(x, y)| dxdy$ vernachlässigbar klein ist.

Dann wird $\sum_{(u,v)\in M} h(u,v) f(x-u,y-v)$ eine gute Approximation für $h \star f(x,y)$ sein, sofern auch f innerhalb eines Quadrates der Seitenlänge 1 nicht sehr schwankt.

Definiere $h_d : \mathbb{Z}^2 \to \mathbb{C}, h_d(x, y) := h(x, y)$ für $(x, y) \in M$ und $h_d(x, y) := 0$ sonst. Dann ist $\mathbb{B}(\mathbb{Z}^2, \mathbb{C}) \to \mathbb{B}(\mathbb{Z}^2, \mathbb{C}), f \mapsto h_d \star f$ ein FIR-Filter, das ähnliche Eigenschaften wie T haben wird.

In praktischen Anwendungen, in denen die Bildsignale nur auf einem beschränkten Rechteck $R \subset \mathbb{Z}^2$ definiert sind, muß man darauf achten, daß das Operatorfenster M klein gegenüber R ist, damit nur in Randnähe Abweichungen auftreten.

In der Praxis muß man eventuell h_d noch normieren, z.B. so daß

$$\sum_{(x,y)\in M} h_d(x,y) = 1 \text{ oder } \sum_{(x,y)\in M} h_d(x,y) = 0$$

gilt, selbst wenn

$$\int_{\mathbb{R}^2} h(x,y) \, dx \, dy = 1 \text{ oder } \int_{\mathbb{R}^2} h(x,y) \, dx \, dy = 0$$

gilt.

Kapitel 8

Anwendungen

8.1 Inverse Filterung

8.1.1 Problemstellung

Ein Bild f werde durch nicht ideale Eigenschaften der Aufnahmeapparatur verändert. Die Veränderung sei als Faltungsoperation $f \mapsto h \star f$ beschreibbar. Zum Beispiel läßt sich eine Defokussierung oder ein Wischeffekt infolge einer Inearen Translationsbewegung annähernd so beschreiben. Aus diesem veränderten Bild $h \star f$ soll das Originalbild rekonstruiert werden.

8.1.2 Lösungsidee

Mache die Faltung rückgängig; man spricht von *Dekonvolution* oder *inverser Filterung*. Ist dies überhaupt möglich? Eine Antwort findet man leicht, wenn man alles im Frequenzbereich betrachtet. Sei $g = h \star f$; dann gilt $\hat{g} = \hat{h} \cdot \hat{f}$ und somit $\hat{f} = \frac{\hat{g}}{\hat{h}}$, also $f = \mathcal{F}^{-1}\left(\frac{\hat{g}}{\hat{h}}\right)$, falls \hat{h} keine Nullstelle hat.

Fazit: Eine Dekonvolution ist möglich, wenn \hat{h} nirgends verschwindet, und zwar in Form eines Filters mit Übertragungsfunktion \hat{h}^{-1} .

8.1.3 Probleme in der Praxis

- 1) Vielfach kennt man h nicht genau.
- 2) Vielfach hat *h* Nullstellen oder nimmt zumindest an einigen Stellen sehr kleine Werte an. Betrachten wir obige Beispiele.
 - a) Defokussierung: lokale Glättung, h ähnlich einer Gaußglocke. Folglich wird $|\dot{h}|$ für hohe Frequenzen sehr klein.
 - b) Verwischung infolge einer Translation: Eine Translation in *x*-Richtung bewirkt idealerweise eine Glättung mit einer Filtermaske $(1, \ldots, 1)$; sie entspricht also einem Boxfilter, dessen Übertragungsfunktion von der Gestalt $\frac{\sin 2a\pi\omega_1}{\pi\omega_1}$ ist, also Nullstellen hat.
- 3) Das Bild $g = h \star f$ liegt nicht mit beliebig hoher Grauwertauflösung vor, sondern ist relativ grob quantisiert (z.B. in nur 256 Graustufen), d.h. man hat gar nicht wirklich g zur Verfügung, sondern nur eine durch Rundungseffekte gestörte Version. Diese Rundungsfehler wirken sich bei der inversen Filterung besonders dort aus, wo $|\hat{h}|$ sehr kleine Werte und folglich $|\hat{h}|^{-1}$ sehr große Werte hat; dort werden die Rundungsfehler unangenehm vergrößert.

Deshalb ist es in der Praxis günstig, $|\hat{h}|^{-1}$ nach oben zu begrenzen oder bei Nullstellen von \hat{h} gleich 0 zu setzen. Dadurch wird die Dekonvolution zwar nicht korrekt, aber immerhin näherungsweise möglich.

4) Die Bildverschlechterung infolge der Faltung mit h hat in Anwendungen häufig Tiefpaßcharakteristik, d.h. $|\hat{h}|$ wird für hohe Frequenzen klein. Und dann ist $|\hat{h}|^{-1}$ für hohe Frequenzen groß, was sehr unangenehm ist, weil man das im Bild f enthaltene Rauschen (z.B. das Quantisierungsrauschen) verstärkt.

Bemerkung (zu 1)). Um h zu bestimmen, kann man folgendes versuchen:

- Paare (f, g) mit $g = h \star f$ messen und daraus \hat{h} bestimmen.
- Falls man die Gestalt von h grob kennt, kann man die genauere Gestalt oft aus den Nullstellen von \hat{h} , die ja auch in \hat{g} vorhanden sind, erraten. Zum Beispiel kann man daran den Radius einer Kreisscheibe ablesen, deren charakteristische Funktion h ist.

8.1.4 Dekonvolution im Diskreten

Sind g und h gegeben, so ist $g = h \star f$ ein lineares Gleichungssystem aus $N_x N_y$ Gleichungen in $N_x N_y$ Unbekannten (den Werten von f). Übliche numerische Lösungsmethoden erfordern einen hohen Aufwand. Durch die DFT wird das Gleichungssystem diagonalisiert; die Koeffizienten auf der Diagonale sind gerade die Werte von \hat{h} .

Sind sie nicht Null, so läßt sich die inverse Matrix trivial berechnen; sie ist einfach die Diagonalmatrix, die die Werte von \hat{h}^{-1} in der Diagonale hat. Da \hat{h} gleich 0 oder zumindest sehr klein werden kann, hat man ein schlecht gestelltes Problem vor sich.

8.1.5 Inverse Filterung verrauschter Signale

Es sei $g = h \star f + r$. Dabei sei r ein Störterm, der als Realisierung eines Rauschprozesses modelliert wird (z.B. könnte man auch die Rundungsfehler beim Quantisieren so modellieren).

Es gilt $\hat{g} = \hat{h} \cdot \hat{f} + \hat{r}$, also $\hat{f} = \frac{\hat{g}}{\hat{h}} - \frac{\hat{r}}{\hat{h}}$, sofern \hat{h} nicht verschwindet. Verglichen mit 8.1.2, ist die Situation noch ungünstiger geworden. Einigermaßen sicher kann man lediglich g bestimmen; \hat{h} kennt man oft schon weniger genau und \hat{r} überhaupt nicht. $\frac{\hat{g}}{\hat{h}}$ wird nur dort eine brauchbare Näherung für \hat{f} sein , wo $|\hat{g}|$ deutlich größer als $|\hat{f}_{\hat{h}}|$ ist. Dies wird nicht überall der Fall sein; denn \hat{r} wird für hohe Frequenzen relativ groß sein, wenn r Realisierung eines weißen Rauschprozesses ist. Deshalb sollte man nur versuchen, eine tiefpaßgefilterte Version von f zu rekonstruieren, also z.B. $\mathcal{F}^{-1}(G \cdot \frac{\hat{g}}{\hat{h}})$, wobei G eine Gaußglocke ist. Aber auch diese Rekonstruktion wird oft sehr schlecht sein. Deshalb bedient man sich oft üblicher Regularisierungsmethoden für inverse, schlecht gestellte Probleme, z.B.

Optimale Constrained Filter: Man geht davon aus, daß man die Varianz V des Rauschen schätzen kann. Gesucht wird dann ein f, so daß $||g - h \star f||_2^2 = V$. Diese Gleichung hat meist viele Lösungen. Man wählt eine aus, die möglichst wenig schwankt, wobei man die Schwankung durch ein Differenzenfilter D mißt (die Auswahl von D erfolgt heuristisch). Man sucht also ein f mit $||g - h \star f||_2^2 = V$, für das $||Df||_2$ minimal ist. Man kann zeigen (siehe z.B. [10]), daß man f aus g durch ein Filter erhält, dessen Übertragungsfunktion aus $\frac{1}{h}$ durch Multiplikation mit einer Funktion entsteht, die die Polstellen von $\frac{1}{h}$ durch Nullstellen kompensiert.

Wiener-Filter: Man betrachtet f und r als Realisierungen stochastischer Prozesse mit Autokorrelationsfunktionen ρ_{ff} und ρ_{rr} und konstruiert ein lineares Filter W, so daß $\mathcal{E}(\|f - Wg\|_2^2)$ minimal ist.

$$\frac{1}{\hat{h}} \cdot \frac{|\hat{h}|^2 \widehat{\rho_{ff}}}{|\hat{h}|^2 \widehat{\rho_{ff}} + \widehat{\rho_{rr}}}$$

ist die Übertragungsfunktion von W; sie hat in den Polstellen von $\frac{1}{\hat{h}}$ Nullstellen. Vergleiche dazu auch stochastische Relaxationsansätze (Geman, Geman).

Meist kennt man ρ_{ff} nicht, und auch ρ_{rr} ist oft nicht exakt bekannt. Man behilft sich dann mit einer Vereinfachung der Übertragungsfunktion von W, und zwar verwendet man

$$H = \frac{1}{\hat{h}} \cdot \frac{|\hat{h}|^2}{|\hat{h}|^2 + \gamma} = \frac{\hat{h}}{|\hat{h}|^2 + \gamma},$$

wobe
i $\gamma>0$ manuell so gewählt wird, daß das restaurierte Bild möglichst kontrastreiche Feinstrukturen ohne offensichtliche Artefakte enthält.

Gegenüber der direkten inversen Filterung (d.h. der Multiplikation mit \hat{h}^{-1}) enthält dieser Ansatz einen zusätzlichen Faktor, der die Polstellen von \hat{h}^{-1} kompensiert; er liefert daher meist deutlich bessere Ergebnisse.

8.2 Korrelationstechniken

8.2.1 Template Matching

Aufgabenstellung: Gegeben sei ein Bild g (eines Objekts oder Musters) und ein größeres Bild f. Finde in f Stellen, die der "Schablone" g stark ähneln.

Lösungsansätze: Seien $f: R \to \mathbb{R}, R = [0, N_x - 1] \times [0, N_y - 1]$ und $g: M \to \mathbb{R}, M = [-n_x, n_x] \times [-n_y, n_y]$, $n_x < N_x, n_y < N_y$.

Es ist üblich, eines der folgenden beiden Kriterien für die Ähnlichkeit von g und f bei Stellen (x, y) mit $(x, y) + M \subset R$ zu wählen.

8.2.1.1 Erster Ansatz

Verwende $\sum_{(u,v)\in M} (g(u,v) - f(x+u,y+v))^2$ als Maß für die Unähnlichkeit. Gesucht werden also Stellen, in denen diese Größe verschwindet oder zumindest ausgeprägte lokale Minima hat.

8.2.1.2 Zweiter Ansatz

Verwende $\sum_{(u,v)\in M} g(u,v)f(x+u,y+v)$ als Maß für die Ähnlichkeit. Gesucht werden also Stellen (x,y), in denen diese Größe lokale Maxima hat.

Begründung: Nach Cauchy-Schwarz gilt

$$\underbrace{\sum_{(u,v)\in M} g(u,v)f(x+u,y+v)}_{=:\langle g,\tau_{(-x,-y)}f\rangle_M} \leq \underbrace{\sqrt{\sum_{(u,v)\in M} g(u,v)^2}}_{=:||g||_M} \cdot \underbrace{\sqrt{\sum_{(u,v)\in M} f(x+u,y+v)^2}}_{=:||\tau_{(-x,-y)}f||_M}$$

wobei $\tau_{(-x,-y)}f(u,v) := f(x+u,y+v).$

Gleichheit gilt bekanntlich genau dann, wenn es ein $\lambda > 0$ mit $g = \lambda \tau_{(-x,-y)} f|_M$ gibt.

Vergleich der beiden Kriterien:

Der erste Ansatz testet auf völlige Übereinstimmung, während der zweite nur auf Übereinstimmung bis auf einen Faktor testet, wodurch unterschiedliche Beleuchtungsstärken keine so große Rolle spielen. Allerdings funktioniert der zweite Ansatz schlecht, wenn sich $\|\tau_{(-x,-y)}f\|_M$ stark mit (x,y) ändert. Man sucht dann besser lokale Maxima des Quotienten $\frac{\langle g, \tau_{(-x,-y)}f \rangle_M}{\|g\|_M \|\tau_{(-x,-y)}f\|_M} \in [-1,1].$

8.2.2 Die Korrelation von Signalen

Für $f, g \in L_1(\mathbb{R}^n)$ oder $f, g \in \mathbb{R}^R$ heißt $\rho_{fg} = f \star g_-$ die Kreuzkorrelation von f mit g, wobei $g_-(p) = g(-p)$ sei. Es gilt $\rho_{fg} \in L_1(\mathbb{R}^n)$ bzw. $\rho_{fg} \in \mathbb{R}^R$ und

$$\rho_{fg}(p) = \int_{\mathbb{R}^n} f(p-q)g(-q)\,dq = \int_{\mathbb{R}^n} f(p+q)g(q)\,dq, \text{ falls } f,g \in L_1(\mathbb{R}^n),$$

bzw.

$$\rho_{fg}(p) = \sum_{q} f(p-q)g(-q) \, dq = \sum_{q} f(p+q)g(q), \text{ falls } f, g \in \mathbb{R}^{R}.$$

Ist f = g, so nennt man ρ_{ff} die Autokorrelation von f.

8.2.2.1 Einige Eigenschaften

a) $\rho_{fg}(p) = \rho_{gf}(-p).$

- b) $\widehat{\rho_{fg}} = \hat{f} \cdot \overline{\hat{g}}$, wenn g reellwertig ist.
- c) $\widehat{\rho_{ff}} = |\widehat{f}|^2$, wenn f reellwertig ist.

Beweis.

a)
$$\rho_{fg}(p) = \int_{\mathbb{R}^n} f(p+q)g(q) \, dq = \int_{\mathbb{R}^n} f(s)g(s-p) \, ds = \rho_{qf}(-p).$$

- b) $\widehat{\rho_{fg}} = \widehat{f \star g_-} = \widehat{f} \cdot (\widehat{g})_- = \widehat{f} \cdot \overline{\widehat{g}}, \text{ weil } g \text{ reell und somit } \overline{\widehat{g}} = (\widehat{g})_- = \widehat{g_-} \text{ gilt.}$
- c) Folgt aus b) mit g = f.

8.2.3 Template Matching und Korrelation

Das Ähnlichkeitsmaß $\sum_{(u,v)\in M} g(u,v) f(x+u,y+v)$ in 8.2.1.2 läßt sich als Kreuzkorrelation $\rho_{fg}(u,v)$ schreiben, wenn man sich g durch den Wert 0 auf R fortgesetzt denkt (oder auch als Faltung mit g_{-}).

Will man Randeffekte infolge der zyklischen Berechnung der Indizes vermeiden, muß man die Bilder durch einen entsprechend breiten Saum von Nullen vergrößern.

Ist M groß, wird der Rechenaufwand im Ortsbereich groß, und es kann vorteilhaft sein, ρ_{fg} vermöge 8.2.2.1.b) über den Frequenzbereich auszurechnen. Dies trifft vor allem für die Berechnung der Autokorrelation ρ_{ff} zu.

Mit Hilfe der Autokorrelation kann man Translationen finden, unter denen sich f selbst ähnelt:

$$\begin{aligned} \|f - \tau_{(-x,-y)}f\|_{2}^{2} &= \|f\|_{2}^{2} + \|\tau_{(-x,-y)}f\|_{2}^{2} - 2\left\langle f, \tau_{(-x,-y)}f\right\rangle = 2\|f\|_{2}^{2} - 2\sum_{(u,v)\in R} f(u,v)f(x+u,y+v) \\ &= 2(\|f\|_{2}^{2} - \rho_{ff}(x,y)) \end{aligned}$$

Lokale Minima von $||f - \tau_{(-x,-y)}f||_2^2$ entsprechen also genau den lokalen Maxima von ρ_{ff} .

8.3 Die Radontransformation

Für $\vartheta \in [0, 2\pi]$ seien $u(\vartheta) = (\cos \vartheta, \sin \vartheta)$ und $u^{\perp}(\vartheta) = (-\sin \vartheta, \cos \vartheta)$. Für $f \in L_1(\mathbb{R}^2)$ und $\vartheta \in [0, \pi[, t \in \mathbb{R} \text{ seien } u(\vartheta) = (-\sin \vartheta, \cos \vartheta)$.

 $f_{\text{Radon}}(\vartheta, t) = \int_{\mathbb{R}} f(tu(\vartheta) + su^{\perp}(\vartheta)) \, ds \qquad \text{die Radontransformierte von } f$ Setze $f_{\vartheta}(t) := f_{\text{Radon}}(\vartheta, t).$

8.3.1 Satz. f ist durch f_{Radon} fast überall eindeutig bestimmt, das heißt, die Radontransformation $f \mapsto f_{\text{Radon}}$ ist injektiv.

Beweis. Weil die Fouriertransformation auf $L_1(\mathbb{R}^2)$ injektiv ist, genügt es zu zeigen, daß \hat{f} durch f_{Radon} eindeutig bestimmt ist. Für jedes $\vartheta \in [0, \pi[$ und $r \in \mathbb{R}$ gilt

$$\begin{split} \hat{f}(ru(\vartheta)) &= \int_{\mathbb{R}^2} f(p) \, e^{-2\pi i \langle ru(\vartheta), p \rangle} \, dp = \int_{\mathbb{R}} \int_{\mathbb{R}} f\left(tu(\vartheta) + su^{\perp}(\vartheta)\right) \, e^{-2\pi i \langle ru(\vartheta), tu(\vartheta) + su^{\perp}(\vartheta) \rangle} \, ds \, dt \\ &= \int_{\mathbb{R}} e^{-2\pi i rt} \int_{\mathbb{R}} f\left(tu(\vartheta) + su^{\perp}(\vartheta)\right) \, ds \, dt = \int_{\mathbb{R}} e^{-2\pi i rt} f_{\vartheta}(t) \, dt = \hat{f}_{\vartheta}(r). \end{split}$$

 u^{\perp} u tu

8.3. DIE RADONTRANSFORMATION

Dieses Ergebnis spielt eine wichtige Rolle und ist bekannt als

8.3.2 Fourier-Slice-Theorem. Für $\vartheta \in [0, \pi[$ und $r \in \mathbb{R}$ gilt $\hat{f}(ru(\vartheta)) = \hat{f}_{\vartheta}(r)$.

8.3.3 Die inverse Radontransformation

Für $p \in \mathbb{R}^2$ und $\hat{f} \in L_1(\mathbb{R}^2)$ gilt

$$\begin{split} f(p) &= \int_{\mathbb{R}^2} \hat{f}(\omega) \, e^{2\pi i \langle p, \omega \rangle} \, d\omega \\ &= \int_0^{2\pi} \int_0^{\infty} \hat{f} \left(ru(\vartheta) \right) \, e^{2\pi i \langle p, ru(\vartheta) \rangle} \, r \, dr \, d\vartheta \\ &= \int_0^{\pi} \int_{\mathbb{R}} \hat{f} \left(ru(\vartheta) \right) \, e^{2\pi i \langle p, ru(\vartheta) \rangle} \, |r| \, dr \, d\vartheta \\ &= \int_0^{\pi} \int_{\mathbb{R}} \hat{f}_{\vartheta}(r) \, e^{2\pi i r \langle p, u(\vartheta) \rangle} \, |r| \, dr \, d\vartheta \\ &= \int_0^{\pi} \left(\mathcal{F}^{-1} \left(|r| \hat{f}_{\vartheta}(r) \right) \right) \left(\langle p, u(\vartheta) \rangle \right) \, d\vartheta \\ &= \int_0^{\pi} \left(\mathcal{F}^{-1} \left(|r| \left(\mathcal{F} f_{\vartheta} \right) (r) \right) \right) \left(\langle p, u(\vartheta) \rangle \right) \, d\vartheta \end{split}$$

Fazit: Die inverse Radontransformation besteht aus der Hintereinanderausführung zweier Operationen:

- 1. $f_{\vartheta} \mapsto \mathcal{F}^{-1}(|r|\mathcal{F}f_{\vartheta})$ Dies ist im Frequenzbereich eine Multiplikation mit der Funktion |r|; im Ortsbereich entspricht ihr eine Faltung $f_{\vartheta} \mapsto \mathcal{F}^{-1}(|r|) \star f_{\vartheta}$, allerdings ist $\mathcal{F}^{-1}(|r|)$ keine Funktion, sondern eine Distribution.
- 2. $g \mapsto \int_0^{\pi} g(\langle p, u(\vartheta) \rangle) d\vartheta$, das heißt, die Anteile aus den einzelnen Richtungen werden aufsummiert.

8.3.4 Typische Anwendung: Röntgentomographie

Eine ebene Scheibe eines Körpers wird in jeder zur Ebene parallelen Richtung ϑ von einer Schar paralleler Röntgenstrahlen in dieser Richtung durchdrungen. Es wird gemessen, wie stark ihre Intensität durch Absorption im Körper geschwächt wird. Ist f ist die Absorptionsdichte der Materie des Körpers, so wird die Abnahme der Intensität der Röntgenstrahlen in Richtung ϑ durch f_{ϑ} gegeben. Die Messung liefert also die Radontransformierte von f. Durch Rücktransformation erhält man die Absorptionsdichteverteilung in der ebenen Scheibe.

8.3.5 Probleme bei der Implementierung

Die zweite Operation ist eigentlich problemlos. Allerdings kann man nur endlich viele Richtungen messen und aufsummieren. Problematisch wird es, wenn man nur Richtungen aus einem eingeschränkten Raumwinkelbereich messen kann.

Die erste Operation ist problematischer. Geht man über den Frequenzbereich, so muß man die Fouriertransformation und deren Inverse auf \mathbb{R} oder \mathbb{Z} durch eine diskrete approximieren (und dabei Daten abschneiden, was problematisch ist, weil die Multiplikation mit |r| eine Hochpaßfilterung ist). Will man im Ortsbereich bleiben, so muß man die Faltung mit der Distribution $\mathcal{F}^{-1}(|r|)$ durch die Faltung mit einer Funktion approximieren, und so eine Funktion hat keinen kleinen Träger. Literaturhinweis: [5].

Kapitel 9

Segmentierung und Klassifikation

Ein allen Stufen der Bildverarbeitung zugrundeliegendes Ziel ist die Segmentierung von Bildern in Bereiche, deren Pixel nach irgendeinem Kriterium zusammengehören.

9.1 Beispielkriterien

- a) Grauwert \leq Schwelle; ergibt: Binärisierung, Färbung von Zusammenhangskomponenten
- b) keine großen Schwankungen der Grauwerte benachbarter Pixel
- c) ähnliche Musterung (*Textur*)
- d) zu einem (eventuell bestimmten) Objekt gehörend; das ist oft nicht mit Mitteln der Bildverarbeitung möglich (Maschinelles Sehen, KI)

9.2 Homogenitätskriterien

Die Segmentierung erfolgt gemäß eines Gleichheits- oder Homogenitätskriterium $H : \mathcal{P}(R) \to \{0, 1\}$, das heißt, A erfüllt $H : \Leftrightarrow H(A) = 1$. Die Auswahl von H erfolgt heuristisch aufgrund der Anwendungssituation. Vielfach entsteht H folgendermaßen:

Auf irgendeine Weise wird zu jedem Pixel p ein Merkmalsvektor $\mu(p) \in \mathbb{R}^n$ bestimmt. Dann gelte H(A) = 1 genau dann, wenn μ innerhalb von A nicht zu sehr schwankt (was noch zu präzisieren ist) bezüglich einer Metrik auf \mathbb{R}^n .

9.3 Beispiele für lokale Merkmalsvektoren

f sei das gegebene Bild.

- a) $\mu = f.$
- b) $\mu = Tf$, wobei T ein Glättungsoperator oder ein Kantenfinder ist.
- c) Lokale Merkmale zur Charakterisierung von Texturen

Es gibt sehr viele Arten, lokale Merkmale zu definieren. Einige typische sind:

- 1) Histogramm einer kleinen Umgebung (aufgefaßt als Vektor im \mathbb{R}^N).
- 2) Autokorrelationsfunktion einer kleinen Umgebung (um Periodizitäten zu erkennen).

- 3) Leistungsspektrum einer kleinen Umgebung (entspricht 2) im Frequenzbereich). Meist wählt man eher die Mittelwerte über Teile des Leistungsspektrums, z.B. wählt man eine Menge von Bandfiltern (eine sogenannte *Filterbank*), deren Durchlaßbereiche den gesamten Frequenzbereich überdecken. Ihre Ausgangsamplituden faßt man zu einem Merkmalsvektor zusammen. Häufig wählt man eine Bank aus Gaborfiltern.
- 4) Statistische Größen der Grauwertverteilung in einer kleinen Umgebung. Meist wählt man Statistiken 1. und 2. Ordnung, also Größen, die von der empirischen Verteilung der Grauwerte (das ist das Histogramm) oder von der empirischen Verteilung von Paaren von Grauwerten abgeleitet sind.

Ein häufiges Hilfsmittel ist die Co-Occurence-Matrix: Für jedes $a \in \mathbb{Z}^2$ und jedes $u, v \in [0, N-1]$ sei

 $\gamma_a(u, v) =$ Anzahl der Pixelpaare (p, q) (in einer vorgegebenen Umgebung \mathcal{U}), für die gilt: q = p + a und f(p) = u und f(q) = v.

Oft normiert man γ_a so, daß $\sum_{(u,v)} \gamma_a(u,v) = 1$, also die $\gamma_a(u,v)$ Häufigkeiten sind.

Die Matrix $(\gamma_a(u, v))_{(u,v)}$ nennt man auch Co-Occurence-Matrix.

Weil eine Texturbeschreibung mit solchen Matrizen für mehrere $a \in \mathbb{Z}^2$ zu umständlich ist, versucht man mit Hilfe dieser Matrizen einige wenige Merkmale zu konstruieren, mit denen die wesentlichen Eigenschaften der Textur erfaßt werden. Einige einfache sind:

- α) Homogenität: $H_1(a) = \sum_{(u,v)} (\gamma_a(u,v))^2$.
- $\beta) \quad \text{Kontrast: } H_2(a) = \sum_{(u,v)} (u-v)^2 \gamma_a(u,v).$
- γ) Entropie: $H_3(a) = -\sum_{(u,v)} \gamma_a(u,v) \log_2 \gamma_a(u,v).$

Aus solchen Größen für einige kleine a bildet man dann einen Merkmalsvektor. Oft vereinfacht man noch dadurch, daß man statt $\gamma_a(u, v)$ mit $a \in \mathbb{Z}^2$ entsprechende richtungsunabhängige Größen definiert wie

 $\gamma_d(u, v) =$ Häufigkeit der Pixelpaare (p, q) mit f(p) = u und f(q) = v und ||p - q|| = d. (d > 0)

Durch vorhergehenden Grauwertausgleich kann man eine gewisse Unabhängigkeit von der Grauwertdynamik erreichen. Man kann auch vorher z.B. einen Kantenoperator anwenden und dann aus dem Ergebnis Co-Occurence-Matrizen bilden.

5) Strukturelle Merkmale. Manche Muster sind mehr oder weniger regelmäßig aus kleinen Elementen, sogenannten *Texeln*, aufgebaut. Man kann dann als Merkmale z.B. Größen verwenden, die die statische Verteilung der Lage und der Orientierung der Texel oder ihre Gestalt charakterisieren.

9.4 Segmentierungsstrategien

Segmentierung in gleichartige Bereiche und Detektion von Grenzen zwischen nicht gleichartigen Bereichen sind in gewissem Sinn komplementär zueinander, aber nicht völlig:

Während man üblicherweise Bereiche sucht, die von geschlossenen Kurven berandet sind, läßt man meist zu, daß Texturkanten irgendwo plötzlich aufhören. Das ist in der Praxis auch oft sinnvoll, z.B. kann der Kontrast längs einer Grauwertkante langsam abnehmen bis auf Null, wo die Kante dann verschwindet. Ebenso können sich Muster benachbarter Bereiche immer ähnlicher werden, bis sie nicht mehr zu unterscheiden sind.

langsame Änderung der Richtung der Texel

Bestimmt man Texturkanten, so erhält man i.a. keine vollständige Segmentierung in Bereiche homogener Texturen. Den beiden Zielrichtungen entsprechen in etwa auch zwei Segmentierungsstrategien.

9.5. CLUSTERANALYSE

9.4.1 Split and Merge

- 1) Wähle eine Anfangszerlegung von R, z.B. in rechteckige Blöcke.
- 2) Verschmelze Nachbarblöcke zu einem größeren Block, wenn dieser das Homogenitätskriterium H erfüllt.
- 3) Jeder Block, der *H* nicht erfüllt, wird nach einer vorgegebenen Regel in kleinere Blöcke aufgeteilt, z.B. geviertelt. Dies wird wiederholt, bis alle Blöcke *H* erfüllen.
- 4) Fasse benachbarte Blöcke zu Bereichen zusammen, wenn dadurch H nicht verletzt wird.

Spezialfall Quadtree: Für jede Teilmenge A eines Bildes setzt man H(A) = 1, wenn alle Pixel in A den gleichen Grauwert haben sind, und H(A) = 0 sonst. Damit wendet man nun die Split-and-Merge an, wobei die Aufteilung stets durch Halbierung der Seitenlängen der Blöcke (so gut es geht) in vier Teilrechtecke erfolgen soll. Die bei dem Zerteilungsvorgang auftretenden Blöcke kann man als Knoten eines Baumes auffassen, in dem von jedem zerteilten Block zu seinen Teilblöcken eine Kante geht. Da somit von jedem Block höchstens vier Kanten ausgehen können, spricht man von einer Quadtree-Darstellung. Sie kann insbesondere für Binärbilder zu Kompressionszwecken verwendet werden.

9.4.2 Region Growing

Bestimme Bereiche, für die das Homogenitätskriterium H seht gut erfüllt ist. Diese Bereiche können recht klein sein und müssen keine Partition des Bildes bilden. Man läßt die Bereiche nun wachsen, indem man immer wieder Pixel aus ihrem Rand hinzunimmt, wenn H — zumindest lokal — nicht verletzt ist. Nach endlich vielen Schritten stoppt dieser Prozeß. Die gewachsenen Bereiche bilden dann nicht unbedingt eine Partition des Bildes. Für Pixel p im Rand zweier Bereiche B_1 und B_2 trifft man folgende Entscheidung: Entweder sind die Merkmale von B_1 und B_2 in einer Umgebung von p gemäß H sehr ähnlich, dann verschmilzt man B_1 und B_2 und nimmt phinzu; oder die Merkmale sind verschieden, dann markiert man p als Kantenpunkt.

9.5 Clusteranalyse

Oft wählt man das Homogenitätskriterium erst in Abhängigkeit von den vorkommenden Daten. Man versucht, die Menge aller im Bild auftretenden Merkmalsvektoren so in endlich viele Teilmengen M_1, \ldots, M_k zu zerlegen, daß die Merkmalsvektoren innerhalb jedes M_j gehäuft liegen und der mittlere Abstand der Vektoren in unterschiedlichen M_j möglichst groß ist (bezüglich einer geeigneten Metrik). Dafür gibt es Verfahren (klassisch und mit neuronalen Netzen). Für $A \subset R$ definiert man dann z.B. H(A) = 1 genau dann, wenn es ein j gibt, so dass alle Merkmalsvektoren der Pixel von A in M_j liegen.

9.6 Klassifikation

Sind Merkmalsklassen M_1, \ldots, M_k vorgegeben, so kann eine Segmentierung des Bildes bezüglich dieser Klassen (eventuell zusätzlich $M_0 = \mathbb{R}^n \setminus \bigcup_{j=1}^k M_j$) erfolgen. Man spricht dann von *Klassifikation*.

Typische Anwendung: Fernerkundung mit Satellitenbildern, Histologie.

9.7 Bemerkungen

Clusteranalyse und Klassifikation beinhalten zwei Aufgaben:

- 1) Lernen der Klassen M_1, \ldots, M_k (mit oder ohne Lehrer).
- 2) Entscheiden, zu welcher Klasse ein vorgelegter Vektor $\in \mathbb{R}^n$ gehört.

Um 2) schnell zu machen, ist es günstig, die M_j durch einfache Ungleichungen zu beschreiben, z.B. polynomiale $\sum_{i=1}^{n} (x_i - z_i)^2 \leq r^2$ (Kugel); die M_j sind dann semialgebraisch. In der Klassifikationstheorie gibt es Verfahren zur Konstruktion solcher Polynomklassifikatoren, die 1) lösen. Eine andere Möglichkeit ist, jedes M_j als Vereinigung von Polyedern (die durch lineare Ungleichungen gegeben sind) darzustellen. Zur Realisierung nimmt man dann neuronale Netze.

Eine Übersicht findet man in [6].

Anhang A

Anwendungsbeispiele

A.1 Richtungsabhängige morphologische Erosion





beschränkte Komponenten

A.2 Double thresholding und feature-AND



Ausgangsbild

die Prufung positiv ausfalle, heist e Seehoters Ministerium, "dann musdie Kassen eben zahlen".

»Ich finde, man darf nicht nur auf einem Auge blind sein.«

Reglerungssprecher Otto Hauser am Mittwoch vergangener Woche im Norddeutschen Rundfunk

Nichtlin. Diff.-Filter mit Totbereich 42

die Prufung positiv ausfalle, heißt e Seehofers Ministerium, "dann muss die Kassen eben zahlen".

»Ich finde, man darf nicht nur auf einem Auge blind sein.«

Regierungssprecher Otto Hauser am M ttwoch vergangener Woche im Norddeutschen Rundfunk die Prüfung positiv ausfalle, heißt e Seehofers Ministerium, "dann müss die Kassen eben zahlen".

Ziel

»lch finde, man darf nicht nur auf einem Auge blind sein.«

Regierungssprecher Otto Hauser am Mittwoch vergangener Woche im Norddeutschen Rundfunk

Nichtlin. Diff.-Filter mit Totbereich 16

die Prufung positiv ausfalle, heußt e Seebofers Ministerium, "dann muss die Kassen eben zahlen".

»Ich finde, man darf nicht nur auf einem Auge blind sein.«

Regierungssprecher Otto Hauser am Mittwoch vergangener Woche im Norddeutschen Rundfunk

Verdünnung durch Skelettierung

88

feature-AND

A.3 Wasserscheide und EDA



Ausgangsbild



Euklidische Distanzabbildung



Resultat mit Trennfehlern

ANHANG A. ANWENDUNGSBEISPIELE

Anhang B

Mathematische Ergänzung

B.1 Metrische Räume, stetige Abbildungen

B.1.1 Definition. (M, d) heißt metrischer Raum, wenn M eine Menge ist und $d : M \times M \to \mathbb{R}$ eine Abbildung mit folgenden Eigenschaften ist:

1) Für alle $x, y \in M$ ist $d(x, y) \ge 0$.

2) d(x,y) = 0 gilt genau dann, wenn x = y.

3) Für alle $x, y \in M$ gilt d(x, y) = d(y, x) (Symmetrie).

4) Für alle $x, y, z \in M$ gilt $d(x, y) \leq d(x, z) + d(z, y)$ (Dreiecksungleichung).

B.1.2 Definition. Sei (M, d) ein metrischer Raum.

a) Als **offene Kugel** B(x,r) mit Mittelpunkt $\in M$ und Radius $a \in [0,\infty)$ bezeichnet die Menge $\{x \in M : d(x,z) < r\}$ und als abgeschlossene Kugel die Menge $\{x \in M : d(x,z) \le r\}$.

b) Eine Teilmenge $U \subset M$ heißt offen, wenn es für jedes $x \in U$ eine Kugel $B(x, r) \subset U$ mit r > 0 gibt. Eine Teilmenge A heißt abgeschlossen, wenn $M \setminus A$ offen ist.

B.1.3 Definition. Ist (M, d) ein metrischer Raum und sind $(x_n)_{n \in \mathbb{N}}$ eine Folge in M und $x \in M$, so heißt die Folge **konvergent** gegen x, wenn es zu jedem $\epsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so dass $d(x_n, x) < \epsilon$ für alle $n \ge n_0$ gilt. x heißt der Grenzwert oder Limes der Folge, kurz $x = \lim_{n \to \infty} x_n$. Er ist eindeutig bestimmt.

B.1.4 Definition. Eine Abbildung $f: M_1 \to M_2$ zwischen zwei metrischen Räumen M_1, d_1) und (M_2, d_2) heißt stetig in $x \in M_1$, wenn für jede Folge $(x_n)_{n \in \mathbb{N}}$ in M_1 , die gegen x konvergiert, gilt

$$f(x) = f(\lim_{n \to \infty} x_n) = \lim_{n \to \infty} f(x_n)$$

f heißt stetig, wenn f in jedem Punkt $x \in M$ stetig ist.

B.1.5 Satz. Für eine Abbildung $f: M_1 \to M_2$ zwischen zwei metrischen Räumen (M_1, d_1) und (M_2, d_2) sind die folgenden Eigenschaften äquivalent:

- (1) f ist stetig.
- (2) Für jede offene Menge $U \in M_2$ ist $f^{-1}(U)$ offen.
- (3) Für jede abgeschlossene Menge $A \in M_2$ ist $f^{-1}(A)$ abgeschlossen.

B.2 Normierte Räume

V sei ein Vektorraum über \mathbb{R} (oder über \mathbb{C}).

B.2.1 Definition. Eine Norm auf V ist eine Abbildung $||||: V \to \mathbb{R}$ mit folgenden Eigenschaften:

1) $||v|| \ge 0$ für jedes $v \in V$.

2) ||v|| = 0 gilt genau dann, wenn v = 0.

3) $\|\alpha v\| = |\alpha| \|v\|$ für alle $v \in V$ und $\alpha \in \mathbb{R}$ (bzw. $\alpha \in \mathbb{C}$).

4) $||v + w|| \le ||v|| + ||w||$ für alle $v, w \in V$ (Dreiecksungleichung).

Anschaulich lässt sich die Norm eines Vektors als Länge deuten.

B.2.2 Beispiel. Normen auf \mathbb{R}^n

Die **p-Norm** im \mathbb{R}^n ist definiert durch

$$||(x_1, \dots, x_n)||_p = \sum_{j=1}^n (|x_j|^p)^{\frac{1}{p}}$$

Dabei ist $p \in \mathbb{R}$ mit $p \ge 1$.

Die 2-Norm

$$||(x_1, \dots, x_n)||_1 = \sqrt{|x_1|^2 + \dots + |x_n|^2}$$

heißt auch **euklidische Norm**; sie entspricht der naiven Vorstellung der Länge eines Vektors. Häufig wird auch die 1-Norm

 $||(x_1,\ldots,x_n)||_1 = |x_1| + \ldots + |x_n|$

verwendet; sie benötigt weniger Rechenaufwand.

Für $p \to \infty$ ergibt sich als Grenzwert der p-Normen die sogenannte ∞ -Norm oder Maximum-Norm

$$||(x_1, \ldots, x_n)||_{\infty} = \max\{|x_1, \ldots, |x_n|\}$$

B.2.3 Beispiel. Normen auf $\mathcal{C}([0,1])$

Auf dem Vektorraum $\mathcal{C}([0,1])$ der stetigen Funktionen $f: [0,1] \to \mathbb{C}$ sind

$$\|f\|_{1} = \int_{0}^{1} |f(t)| dt \quad \text{und} \quad \|f\|_{2} = \int_{0}^{1} |f(t)|^{2} dt \quad \text{und} \quad \|f\|_{\infty} = \max\{|f(t)| : t \in [0,1]\}$$

Normen.

B.2.4 Satz. Ist |||| eine Norm, so wird durch d(x, y) = ||x - y|| eine Metrik definiert. **Bemerkung:** Nicht jede Metrik auf einem Vektorraum wird auf diese Weise von einer Norm induziert!

B.3 Euklidische Vektorräume, Skalarprodukt

V sei ein Vektorraum über \mathbb{R} .

B.3.1 Definition. Ein **Skalarprodukt** auf V ist eine Abbildung $\langle | \rangle : V \times V \to \mathbb{R}$ mit folgenden Eigenschaften:

1) $\langle x | x \rangle \geq 0$ für jedes $x \in V$.

 $2) < x|y > = < y|x > \text{für alle } x, y \in V.$

3) < | > ist bilinear d.h. für alle $x,y,z\in V$ und alle $\alpha,\beta\in\mathbb{R}$ gilt

$$< \alpha x + \beta y | z > = \alpha < x | z > + \beta < y | z >$$

und

$$\langle z | \alpha x + \beta y \rangle = \alpha \langle z | x \rangle + \beta \langle z | y \rangle$$

Ein Vektorraum mit Skalarprodukt heißt euklidischer Vektorraum.

B.3.2 Beispiel. Skalarprodukte auf \mathbb{R}^n

Für jede symmetrische, positiv definite Matrix A wird durch

$$\langle x|y \rangle = xAy^{\top} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_jy_j$$
 für $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n) \in \mathbb{R}^n$

ein Skalarprodukt auf dem \mathbb{R}^n definiert. Wählt man für A die Einheitsmatrix, so erhält man das euklidische Skalarprodukt

$$\langle x|y \rangle = x \cdot y^{\top} = \sum_{j=1}^{n} x_j y_j$$
 für $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n) \in \mathbb{R}^n$.

B.3.3 Beispiel. Skalarprodukt auf C([0,1])

Auf dem Vektorraum $\mathcal{C}([0,1])$ der stetigen Funktionen $f: [0,1] \to \mathbb{R}$ ist

$$< f|g> = \int_0^1 f(t)g(t)dt$$

ein Skalarprodukt.

B.3.4 Satz. Ist < | > ein Skalarprodukt, so wird durch $||x|| = \sqrt{\langle x|x \rangle}$ eine Norm definiert. Für alle $x, y \in V$ gilt dann die **Cauchy-Ungleichung**

$$< x|y > | \le ||x|| \cdot ||y|$$

Gleichheit gilt genau dann, wenn x und y linear abhängig sind.

Aus der Cauchy-Ungleichung folgt, dass für all
e $x,y\in V\setminus\{0\}$ stets

$$-1 \le \frac{< x | y >}{\|x\| \|y\|} \le 1$$

Daher existiert $\phi := \arccos \frac{\langle x | y \rangle}{\|x\| \|y\|}$, und damit gilt

$$\langle x|y\rangle = \|x\|\|y\|\cos\phi$$

Der Wert ϕ wird der Winkel zwischen x und y genannt; er ist nur definiert, wenn x und y beide ungleich Null sind. Wenn $\langle x|y \rangle = 0$ ist, so sagt man, x und y stehen **senkrecht** aufeinander. Sofern sie nicht 0 sind, ist dies gleichbedeutend damit, dass der Winkel zwischen ihnen $\pm 90^{\circ}$ beträgt.

B.3.5 Bemerkung. Nicht jede Norm wird auf obige Weise durch ein Skalarprodukt erzeugt. Notwendig und hinreichend dafür ist die Gültigkeit der Parallelogrammidentität

$$||x+y||^2 + ||x-y||^2 = 2||x||^2 + 2||y||^2$$
 für alle x, y .

B.4 Unitäre Vektorräume, Hermitesches Skalarprodukt

Einen Vektorraum über \mathbb{C} kann man natürlich als Vektorraum über \mathbb{R} ansehen und dann obige Überlegungen anwenden. Man kann aber auch seine komplexe Struktur explizit ausnutzen und dementsprechend den Begriff des Skalarproduktes zu dem einer hermiteschen Form erweitern.

V sei ein Vektorraum über \mathbb{C} .

B.4.1 Definition. Ein hermitesches Skalarprodukt auf V ist eine Abbildung $\langle | \rangle : V \times V \to \mathbb{R}$ mit folgenden Eigenschaften:

1) $\langle x | x \rangle \geq 0$ für jedes $x \in V$.

 $2) < x|y >= \overline{\langle y|x \rangle} \text{ für alle } x, y \in V.$

3) < | > ist sesquilinear d.h. für all
e $x,y,z\in V$ und alle $\alpha,\beta\in\mathbb{C}$ gilt

$$< \alpha x + \beta y | z >= \alpha < x | z > + \beta < y | z >$$

und

 $< z | \alpha x + \beta y > = \overline{\alpha} < z | x > + \overline{\beta} < z | y >$

Ein Vektorraum mit Skalarprodukt heißt unitärer Vektorraum.

B.4.2 Beispiel. Hermitesche Skalarprodukte auf \mathbb{C}^n

Für jede hermitesche, positiv definite Matrix A wird durch

$$\langle x|y \rangle = xA\overline{y}^{\top} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_j\overline{y_j}$$
 für $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n) \in \mathbb{R}^n$

eine hermitesches Skalarprodukt definiert. Wählt man für A die Einheitsmatrix, so erhält man das schon oben definierte **euklidische Skalarprodukt**.

B.4.3 Beispiel. Hermitesches Skalarprodukt auf $\mathcal{C}([0,1])$

Auf dem Vektorraum $\mathcal{C}([0,1])$ der stetigen Funktionen $f: [0,1] \to \mathbb{C}$ ist

$$< f|g> = \int_0^1 f(t) \overline{g(t)} dt$$

ein hermitesches Skalarprodukt.

B.4.4 Satz. Ist $\langle | \rangle$ ein hermitesches Skalarprodukt, so wird durch $||x|| = \sqrt{\langle x|x\rangle}$ eine Norm definiert. Für alle $x, y \in V$ gilt dann die **Cauchy-Ungleichung**

$$|\langle x|y \rangle| \le ||x|| \cdot ||y||$$

Gleichheit gilt genau dann, wenn x und y linear abhängig sind. Winkel lassen sich damit nicht sinnvoll definieren.

B.5 Orthonormalbasen

Sei V ein endlichdimensionaler, euklidischer oder unitärer Vektorraum.

B.5.1 Definition. Eine Basis b_1, \ldots, b_n von V heißt **Orthonormalbasis** (kurz ONB), wenn gilt:

1) $||b_j|| = 1$ für alle j = 1, ..., n

2) $\langle b_j | b_k \rangle = 0$ für alle j, k mit $j \neq k$

B.5.2 Satz. Ist b_1, \ldots, b_n eine Orthonormalbasis von V, so gilt für jedes $x \in V$

$$x = \sum_{j=1}^{n} \langle x | b_j \rangle > b_j$$

Die Entwicklungskoeffizienten bezüglich einer ONB lassen sich also sehr viel leichter berechnen als bezüglich einer beliebigen Basis.

B.6 Maßräume, Wahrscheinlichkeitsmaße

B.6.1 Definition. Ein Messraum ist ein Paar (M, \mathcal{A}) , wobei M eine Menge ist und \mathcal{A} eine σ -Algebra über M ist, die durch folgende Eigenschaften definiert wird:

1) \mathcal{A} ist eine Teilmenge der Potenzmenge von M.

2) $M \in \mathcal{A}$.

3) Abgeschlossenheit unter Komplementbildung: Ist $A \in \mathcal{A}$, so ist auch $M \setminus A \in \mathcal{A}$.

4) Abgeschlossenheit unter abzählbaren Vereinigungen:

Ist $(A_n)_{n\in\mathbb{N}}$ eine abzählbare Folge von Mengen $A_n \in \mathcal{A}$, so gilt $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Die Elemente aus \mathcal{A} nennt man die **messbaren Mengen** des Messraumes.

B.6.2 Beispiel. Für jede Menge M ist die Potenzmenge von M eine σ -Algebra..

B.6.3 Beispiel. Die Borelschen Mengen im \mathbb{R}^n

Seien $M = \mathbb{R}^n$ und \mathcal{A} die kleinste σ -Algebra über \mathbb{R}^n , die alle Quader enthält. Diese σ -Algebra heißt die σ -Algebra der Borelschen Mengen \mathcal{B}_n .

Sie existiert, denn sie ist einfach der Durchschnitt derjenigen σ -Algebren, welche alle Quader enthalten, und solche σ -Algebren gibt es, z.B. die Potenzmenge $\mathcal{P}(\mathbb{R}^n)$. Man kann zeigen, dass \mathcal{B}_n echt kleiner als $\mathcal{P}(\mathbb{R}^n)$ ist, dass es also Teilmengen von \mathbb{R}^n gibt, die nicht Borelsch sind.

Ob man bei der Definition von \mathcal{B}_n alle Quader zulässt oder nur die offenen oder nur die abgeschlossenen, spielt keine Rolle. Jede abzählbare Menge ist Borelsch.

B.6.4 Definition. Ein **Maß** μ auf einem Messraum (M, \mathcal{A}) ist eine Abbildung $\mu \colon \mathcal{A} \to [0, \infty]$ mit folgenden Eigenschaften:

1) $\mu(\emptyset) = 0.$

2) Für jede Folge $(A_j)_{j\in\mathbb{N}}$ in \mathcal{A} aus paarweise disjunkten Mengen A_j gilt

$$\mu(\bigcup_{j\in\mathbb{N}}A_j) = \sum_{j\in\mathbb{N}}\mu(A_j) \qquad (\sigma\text{-Additivität})$$

Man nennt das Tripel (M, \mathcal{A}, μ) einen **Maßraum**.

Anschaulich kann man ein Maß μ als eine Vorschrift deuten, die jeder messbaren Menge eine Art Volumen zuordnet. Aus 1) und 2) folgt sofort, dass stets $\mu(A \cup B) \leq \mu(A) + \mu(B)$ gilt und dass außerdem $\mu(A) \leq \mu(B)$ für $A \subset B$. Eine Menge mit Maß 0 wird **Nullmenge** genannt. Jede messbare Teilmenge einer Nullmenge ist ebenfalls eine Nullmenge.

B.6.5 Definition. Ein Maßraum (M, \mathcal{A}, μ) heißt ein **Wahrscheinlichkeitsraum**, wenn $\mu(M) = 1$ gilt. Das Maß μ wird dann ein **Wahrscheinlichkeitsmaß** genannt; sein Wertebereich ist [0, 1]. Häufig bezeichnet man Wahrscheinlichkeitsmaße mit dem Buchstaben p (für probability).

Anschaulich besteht sind die Elemente von M die möglichen Ergebnisse eines Experimentes, und die Elemente von \mathcal{A} sind diejenigen Mengen A von Ergebnissen, denen man einen Wahrscheinlichkeitswert zuordnen kann, nämlich p(A).

B.6.6 Beispiel. Endliche Wahrscheinlichkeitsräume

Ein Experiment, das nur endlich viele Ergebnisse $x_1, \ldots x_n$ haben kann, beschreibt man durch die Ergebnismenge $M = \{x_1, \ldots, x_n\}$, die σ -Algebra $\mathcal{A} = \mathcal{P}(M)$ und die Wahrscheinlichkeiten $p(\{x_1\}), \ldots, p(\{x_n\})$ der einelementigen Teilmengen von M, die sich zu 1 summieren. Setzt man $p(A) = \sum_{a \in A} p(\{a\})$ für eine beliebige Teilmenge $A \subset M$, so erhält man eine Abbildung $p: \mathcal{A} \to [0, 1]$, die die Maßeigenschaften 1) und 2) erfüllt.

B.6.7 Beispiel. Das Lebesgue-Maß auf \mathbb{R}^n

Für jeden Quader $Q = [a_1, b_1] \times \ldots \times [a_n, b_n]$ im \mathbb{R}^n sei

$$\mu(Q) = (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n)$$

das elementargeometrische Volumen des Quaders. Man kann zeigen, dass μ auf eindeutige Weise zu einem Maß $\lambda \colon \mathcal{B}_n \to [0, \infty]$ auf die Borelschen Mengen fortgesetzt werden kann. Dieses Maß heißt das **Lebesgue-Maß**. Auf den Borelschen Teilmengen des Einheitsquaders wird dadurch ein Wahrscheinlichkeitsmaß induziert.

Jede einpunktige Menge im \mathbb{R}^n hat das Lebesguemaß 0; folglich ist auch jede abzählbare Menge eine **Lebesgue-Nullmenge**, insbesondere ist die Menge \mathbb{Q} der rationalen Zahlen eine Lebesgue-Nullmenge.

Bemerkung: Ist μ ein Maß auf den Borelschen Mengen \mathcal{B}_n , so muss nicht jede Teilmenge einer Menge $A \in \mathcal{B}_n$ mit $\mu(A) = 0$ wieder in \mathcal{B}_n sein. Man kann aber das Maß μ auf eine größere, \mathcal{B}_n umfassende σ -Algebra \mathcal{A} fortsetzen, so dass jede Teilmenge einer Nullmenge von \mathcal{A} wieder in \mathcal{A} liegt (und natürlich auch eine Nullmenge ist). Man spricht dann von der Vervollständigung des Maßes μ . Grob gesagt entsteht \mathcal{A} durch die Hinzunahme der Teilmengen von μ -Nullmengen aus \mathcal{B}_n . Infolgedessen hängt \mathcal{A} von dem Maß μ ab.

B.7 Messbare Abbildungen, Zufallsvariablen, Verteilungen, Dichten

B.7.1 Definition. Seien (M_1, \mathcal{A}_1) und (M_2, \mathcal{A}_2) Messräume. Eine Abbildung $f: M_1 \to M_2$ heißt messbar, wenn $f^{-1}(A) \in \mathcal{A}_1$ für jedes $A \in \mathcal{A}_2$.

Das Urbild einer messbaren Menge unter einer messbaren Abbildung ist also wieder messbar (man vergleiche die Analogie zur Stetigkeit). Deshalb kann man mit einer messbaren Abbildung Maße auf dem Definitionsbereich der Abbildung in den Bildraum transportieren.

B.7.2 Satz. Sind $(M_1, \mathcal{A}_1, \mu)$ ein Maßraum, (M_2, \mathcal{A}_2) ein Messraum und $f: M_1 \to M_2$ eine messbare Abbildung, so ist $\mu_*(A): = \mu \circ f^{-1}$ ein Maß auf (M_2, \mathcal{A}_2) ; es gilt

$$\mu_*(A) = \mu(f^{-1}(A))$$
 für jedes $A \in \mathcal{A}_2$.

Dieses Maß μ_* heißt das **Bildmaß von** μ **unter** f.

Im Falle, dass μ ein Wahrscheinlichkeitsmaß ist, wird eine messbare Abbildung f meist eine **Zufallsvariable** genannt und das Bildmaß μ_* wird die **Verteilung** von f genannt. Zufallsvariablen werden gern mit Großbuchstaben wie X oder Y bezeichnet.

In Anwendungen kommen meist Zufallsvariablen vor, deren Wertebereich \mathbb{R}^n ist; dies liegt daran, dass die Messergebnisse von Experimenten meist numerische Größen sind. Den Wahrscheinlichkeitsraum (M, \mathcal{A}, p) , der dem gesamten Experiment zugrunde liegt, kennt man oft nicht explizit. Man kann lediglich beobachten, welche Werte einige Zufallsvariablen bei Ausführung des Experimentes annehmen. Deshalb spielen die Verteilungen der Zufallsvariablen oft eine größere Rolle als das Wahrscheinlichkeitsmaß p. Diese Verteilungen kann man empirisch zu schätzen versuchen, oder man kann aufgrund von Vorwissen über das Experiment gewisse Annahmen über die Gestalt der Verteilungen machen. Bei numerischen Zufallsvariablen, deren Bildraum also der \mathbb{R}^n mit den Borelschen Mengen \mathcal{B}_n als σ -Algebra ist, nimmt man häufig an, dass ihre Verteilung eine Dichte bezüglich des Lebesguemaßes auf \mathbb{R}^n hat.

B.7.3 Definition. Ein Maß μ auf $(\mathbb{R}^n, \mathcal{B}_n)$ hat eine **Dichte** h (bezüglich des Lebesguemaßes), wenn es eine (Lebesgue-)integrierbare Funktion $h: \mathbb{R}^n \to [0, \infty[$ gibt, so dass

$$\mu(A) = \int_A h(x) dx \quad \text{ für jedes } A \in \mathcal{B}_n$$

gilt.

Man kann sich überlegen, dass h nur bis auf eine Nullmenge bestimmt ist. In vielen Anwendungen ist h stetig oder zumindest stückweise stetig, so dass man das Integral auch als eigentliches oder uneigentliches Riemann-Integral auffassen kann, wenn A ein Intervall ist.

B.7.4 Beispiel. Die Gleichverteilung auf einem Intervall.

Seien $a, b \in \mathbb{R}, a < b$ und M = [a, b] ein Intervall in \mathbb{R} . Dann ist die Funktion $h: [a, b] \to \mathbb{R}$ mit $h(t) = \frac{1}{b-a}$ die Dichte eines Wahrscheinlichkeitsmaßes p auf [a, b]. Dieses Maß p heißt die Gleichverteilung auf [a, b], weil die Dichte konstant ist und somit die Wahrscheinlichkeit jeder Borelschen Menge proportional zu ihrem Lebesguemaß ist.

B.7.5 Beispiel. Die Normal- oder Gaußverteilung auf \mathbb{R} .

Für $\alpha, \sigma \in \mathbb{R}$ und $\sigma > 0$ nennt man die Funktion

$$h\colon \mathbb{R}\to \mathbb{R}, \quad h(t)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\alpha)^2}{2\sigma^2}}$$

eine Gaußfunktion. Der Normierungsfaktor ist so gewählt, dass

$$\int_{\mathbb{R}} h(t)dt = 1$$

gilt. *h* ist also die Dichte eines Wahrscheinlichkeitsmaßes auf \mathbb{R} , das man eine **Gaußverteilung oder Nor**malverteilung nennt und mit $\mathcal{N}(\alpha, \sigma^2)$ bezeichnet.

Die Gaußverteilung wird bei der Modellbildung in vielen Anwendungen verwendet; das liegt nicht nur daran, dass sie analytisch sehr schöne Eigenschaften hat, sondern auch daran, dass sie infolge des zentralen Grenzwertsatzes häufig eine gute Näherung an die tatsächliche Verteilung einer Zufallsvariable ist, die oft nicht exakt bekannt ist.

Manchmal nennt man auch die Funktion ohne den Normierungsfaktor eine Gaußfunktion. Wir bezeichnen sie mit

$$g_{\sigma}(t) = e^{-\frac{t^2}{2\sigma^2}}$$

B.7.6 Beispiel. Die mehrdimensionale Normal- oder Gaußverteilung.

Für $\alpha_1, \ldots, \alpha_n, \sigma_1, \ldots, \sigma_n \in \mathbb{R}$ und $\sigma_j > 0$ für alle j nennt man die Funktion

$$h: \mathbb{R}^n \to \mathbb{R}, \quad h(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n} \prod_{j=1}^n \sigma_j} \exp\left(-\sum_{j=1}^n \frac{(x_j - \alpha)^2}{2\sigma_j^2}\right)$$

eine **n-dimensionale Gaußfunktion**. Auch sie ist die Dichte eines Wahrscheinlichkeitsmaßes auf \mathbb{R}^n , das man die n-dimensionale Normal- oder Gaußverteilung nennt.

Haben alle σ_i den gleichen Wert σ , so spricht man von einer isotropen Gaußfunktion

$$h: \mathbb{R}^n \to \mathbb{R}, \quad h(x_1, \dots, x_n) = \frac{1}{\sigma^n \sqrt{(2\pi)^n}} \exp\left(-\frac{\sum_{j=1}^n (x_j - \alpha_j)^2}{2\sigma^2}\right) = \frac{1}{\sigma^n \sqrt{(2\pi)^n}} \exp\left(-\frac{\|x - \alpha\|_2^2}{2\sigma^2}\right)$$

Dabei sind $x = (x_1, \ldots, x_n)$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ und $|| ||_2$ die euklidische Norm. Bezeichnet man mit $g_{\sigma}(t) = \exp(-\frac{t^2}{2\sigma^2})$ die nicht normierte Gaußfunktion in einer Dimension, so hat die n-dimensionale Gaußfunktion die Gestalt

$$h(x) = \frac{1}{\sigma^n \sqrt{(2\pi)^n}} g_\sigma(\|x - \alpha\|_2)$$

B.8 Integrale

Es sei hier nur sehr informell an die Unterschiede von Riemann- und Lebesgue-Integral erinnert.

Das eindimensionale Riemann-Integral

Es seien [a, b] ein endliches Intervall in \mathbb{R} und $f : [a, b] \to \mathbb{R}$ eine beschränkte Funktion. Eine Zerlegung \mathcal{Z} von [a, b] ist eine endliche, streng monoton wachsende Folge $x_0 = a < x_1 < \ldots < x_{k-1} < x_k = b$. Der maximale Abstand zweier aufeinanderfolgender Folgenglieder, also max $\{x_j - x_{j-1} : j = 1, \ldots, k\}$, heißt die Feinheit der Zerlegung.

Bezüglich so einer Zerlegung bildet man die

die Riemannsche Untersumme
$$\underline{S}_{\mathcal{Z}}f = \sum_{j=1}^{k} m_j(x_j - x_{j-1})$$

und

die Riemannsche Obersumme
$$\overline{S}_{\mathcal{Z}}f = \sum_{j=1}^{k} M_j(x_j - x_{j-1})$$

wobei $m_j = \min\{f(t) : x_{j-1} \le t \le x_j\}$ und $M_j = \max\{f(t) : x_{j-1} \le t \le x_j\}$ sind.

Konvergieren die Ober- und die Untersummen, wenn die Feinheiten der Zerlegungen gegen 0 gehen, gegen denselben Grenzwert, so nennt man die Funktion f Riemann-integrierbar und den Grenzwert das Riemann-Integral der Funktion f und bezeichnet es mit $\int_a^b f(x) dx$.

Die Riemann-integrierbaren Funktion bilden einen Vektorraum über \mathbb{R} . Dazu gehören alle stückweise stetigen Funktionen. Und das Riemann-Integral ist ein lineares Funktional darauf.

Für Funktionen f, die auf ganz \mathbb{R} definiert sind, definiert man

das uneigentliche Riemann-Integral
$$\int_{-\infty}^{\infty} f(x) dx = \lim_{n \to \infty} \int_{a_n}^{b_n} f(x) dx$$

wenn der Limes für alle gegen $-\infty$ konvergierenden Folgen a_n und alle gegen ∞ konvergierenden Folgen b_n existiert, und

den Cauchyschen Hauptwert
$$CHf = \lim_{n \to \infty} \int_{-a_n}^{a_n} f(x) dx$$

wenn der Limes für alle gegen ∞ konvergierenden Folgen a_n existiert.

Leider hat das Riemann-Integral nicht genügend gute Eigenschaften, um effizient Analysis treiben zu können. Schon manche einfache Funktionen wie z.B. die charakteristische Funktion der rationalen Zahlen in [0, 1] sind nicht Riemann-integrierbar (weil die Obersummen alle gleich 1 und die Untersummen alle gleich 0 sind). Und es gibt keine guten Sätze über die Integrabilität der Grenzwerte von Funktionenfolgen sowie über die Vertauschbarkeit von Integral und Grenzwerten.

Deshalb wurden andere Integralbegriffe konstruiert; der wohl bekannteste ist der von Lebesgue. Wir wollen hier nur kurz eine mögliche Einführung skizzieren, welche insbesondere den Unterschied zum Riemannintegral verdeutlicht.



Das Lebesgue-Integral

Es sei $f: \mathbb{R} \to \mathbb{R}$ eine Borel-messbare Funktion, λ das Lebesguemaß auf \mathbb{R} . Man betrachtet nun nicht wie bei dem Riemann-Integral Zerlegungen des Definitionsbereichs, sondern des Bildbereichs. So eine Zerlegung \mathcal{Z} ist eine streng monotone wachsende Folge $(z_j)_{j\in\mathbb{Z}}$ in \mathbb{R} . Das Supremuum der Abstände aufeinanderfolgender Folgenglieder wird wieder die Feinheit genannt. Für jede solche Zerlegung und jede Folge $\alpha = (\alpha_j)_{j\in\mathbb{Z}}$ mit $z_j \leq \alpha_j \leq z_{j+1}$ definiert man

die Lebesgue-Summe
$$S_{\mathcal{Z},\alpha}f = \sum_{j\in\mathbb{Z}} \alpha_j \lambda \left(f^{-1}\left([z_j, z_{j+1}] \right) \right)$$

Diese Summen müssen nicht konvergieren. Wenn jedoch eine Lebesguesumme absolut konvergiert, so auch alle anderen und ihre Werte konvergieren mit fallender Feinheit der Zerlegung gegen einen Grenzwert, den man das **Lebesgue-Integral** von f nennt. Unmittelbare Folgerung dieser Definition ist, dass f genau Lebesgue-integrierbar ist, wenn |f| es ist.

Die Urbilder $f^{-1}([z_j, z_{j+1}])$ der Intervalle der Zerlegung sind nicht notwendigerweise wieder Intervalle, sondern irgendwelche Borelsche Mengen. Daher braucht man also den Begriff des Lebesguemaßes, um Lebesguesummen überhaupt hinschreiben zu können.

Dieser Ansatz zur Einführung des Lebesgueintegrals lässt sich verallgemeinern, indem man einen beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, p) als Definitionsbereich zulässt. Eine Lebesguesumme hat dann die Gestalt

$$S_{(\mathcal{Z},\alpha)}F = \sum_{j\in\mathbb{Z}} \alpha_j p\left(f^{-1}\left([z_j, z_{j+1}[]\right)\right)$$

und das Integral von f über Ω wird mit $\int_{\Omega} f(\omega) dp(\omega)$ oder $\int_{\Omega} f dp$ bezeichnet.

Bemerkungen

1. Die charakteristische Funktion der Menge der rationalen Zahlen in einem beschränkten Intervall ist nicht Riemann-integrierbar, aber sehr wohl Lebesgue-integrierbar, weil die rationalen Zahlen eine Lebesgue-Nullmenge sind.

2. Jede Riemann-integrierbare Funktion auf einem beschränkten Intervall ist auch Lebesgue-integrierbar.

3. Nicht jede uneigentlich Riemann-integrierbare Funktion ist Lebesgue-integrierbar.

Gegenbeispiel: $f(x) = \frac{\sin x}{x}$ für $x \neq 0$ und f(0) = 1. Diese Funktion spielt in der Signalverarbeitung eine wichtige Rolle. Wäre sie Lebesgue-integrierbar, so wäre es auch ihr Betrag |f|; das kann aber nicht sein, weil die Fläche zwischen dem Graphen von |f| und der x-Achse nicht endlich ist (sie kann durch die harmonische Reihe nach unten abgeschätzt werden).

Die große Bedeutung des Lebesgue-Integrals besteht darin, dass es für das Lebesgue-Integral gut anwendbare Konvergenzsätze gibt, also Aussagen darüber, wann der Grenzwert einer Folge integrierbarer Funktionen wieder integrierbar ist und ihre Integrale gegen das Integral des Grenzwertes konvergieren. Insbesondere erhält man damit auch Aussagen, wann Differentiation und Integral vertauscht werden dürfen.

B.9 Produkträume, stochastische Unabhängigkeit

Endliche Produkte

Seien $(\Omega_j, \mathcal{A}_j, p_j), j = 1, \ldots, n$ Wahrscheinlichkeitsräume. Dann definiert man auf dem kartesischen Produkt $\Omega = \Omega_1 \times \ldots \times \Omega_n$ die Produkt- σ -Algebra \mathcal{A} als kleinste σ -Algebra, die alle sogenannten Zylindermengen der Gestalt $A_1 \times \ldots \times A_n$ mit $A_j \in \mathcal{A}_j$ enthält. Auf diesen Zylindermengen setzt man

$$p(A_1 \times \ldots \times A_n) = p_1(A_1) \cdot \ldots \cdot p_n(A_n)$$

Man kann zeigen, dass dieses p eindeutig zu einem Maß p auf der Produkt- σ -Algebra \mathcal{A} fortgesetzt werden kann. (Ω, \mathcal{A}, p) heißt das Produkt der Wahrscheinlichkeitsräume $(\Omega_j, \mathcal{A}_j, p_j)$ und p das Produkt der Maße p_j , in Zeichen $p = p_1 \otimes \ldots \otimes p_n$.

B.9.1 Satz (Fubini).

Ist (Ω, \mathcal{A}, p) das Produkt der Wahrscheinlichkeitsräume $(\Omega_1, \mathcal{A}_1, p_1)$ und $(\Omega_2, \mathcal{A}_2, p_2)$ und ist $f: \Omega \to \mathbb{R}$ eine \mathcal{A} -messbare Funktion, so gilt

$$\int_{\Omega} f \, dp = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, dp_2(\omega_2) \, dp_1(\omega_1) = \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \, dp_1(\omega_1) \, dp_2(\omega_2),$$

insbesondere existieren alle Integrale.

Bemerkungen: Diese Definition lässt sich unmittelbar auf Maßräume übertragen, deren Maß nicht endlich, aber σ -endlich ist. Das trifft z.B. für das Lebesguemaß auf \mathbb{R} zu. Das Lebesguemaß auf \mathbb{R}^n ist in diesem Sinne das Produkt von n eindimensionalen Lebesguemaßen.

Seien (Ω, \mathcal{A}, p) ein Wahrscheinlichkeitsraum und $X_j: \Omega \to \mathbb{R}, j = 1, \ldots, n$, Zufallsvariablen. Setze $X: \Omega \to \mathbb{R}^n, \omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$. $\mu_j = p \circ X_j^{-1}$ sei die Verteilung von X_j . Die Verteilung $\mu = p \circ X^{-1}$ von X heißt die gemeinsame Verteilung der X_j .

B.9.2 Definition.

Die Zufallsvariablen X_1, \ldots, X_n heißen **stochastisch unabhängig**, wenn ihre gemeinsame Verteilung das Produktmaß der einzelnen Verteilungen ist, wenn also $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ gilt.

Viele Sätze der Stochastik machen Aussagen darüber, was passiert, wenn man ein Experiment beliebig oft wiederholt, z.B. das starke und das schwache Gesetz der großen Zahlen. Um solche Aussagen überhaupt erst einmal exakt formulieren zu können, braucht man ein mathematisches Modell für ein unendlich oft wiederholtes Experiment. Dazu benötigt man das Produkt unendlich vieler Wahrscheinlichkeitsräume.

Unendliche Produkte

I sei eine beliebige Indexmenge. Für jedes $i \in I$ sei ein Wahrscheinlichkeitsraum $(\Omega_i, \mathcal{A}_i, p_i)$ gegeben. Auf dem kartesischen Produkt $\prod_{i \in I} \Omega_i$ sei $\bigotimes_{i \in I} \mathcal{A}_i$ die kleinste σ -Algebra, die alle endlichen Zylindermengen d.h. alle Mengen der Gestalt $\prod_{j \in J} A_j \times \prod_{i \in I \setminus J} \Omega_i$ enthält, für die J eine endliche Teilmenge von I ist und $A_j \in \mathcal{A}_j$ für jedes $j \in J$ gilt. Diese σ -Algebra \mathcal{A} heißt das **Produkt der** σ -Algebren \mathcal{A}_i . Sie ist die kleinste σ -Algebra, bezüglich der alle Projektionen $pr_j : \prod_{i \in I} \Omega_i \to \Omega_j$ messbar sind.

Für jede Zylindermenge $\prod_{j\in J}A_j\times\prod_{i\in I\setminus J}\Omega_i$ definiert man

$$\mu\bigg(\prod_{j\in J} A_j \times \prod_{i\in I\setminus J} \Omega_i\bigg) = \prod_{j\in J} p_j(A_j)$$

Diese Abbildung lässt sich eindeutig zu einem Wahrscheinlichkeitsmaß μ auf $\bigotimes_{i \in I} \mathcal{A}_i$ fortsetzen; man nennt μ das **Produktmaß der** p_i und bezeichnet es mit $\bigotimes_{i \in I} p_i$.

Wird ein Experiment durch einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, p) beschrieben, so modelliert man die unendliche, unabhängige Wiederholung des Experiments als Produktraum der abzählbaren Familie $(\Omega_i, \mathcal{A}_i, p_i)_{i \in \mathbb{N}}$, bei der alle Glieder gleich (Ω, \mathcal{A}, p) sind.

Ist $X: \Omega \to \mathbb{R}$ eine Zufallsvariable, so werden durch $X_i = X \circ pr_i$ unabhängige Zufallsvariablen $X_i: \Omega^{\mathbb{N}} \to \mathbb{R}$ definiert. Anschaulich ist X_i die Zufallsvariable X bei der *i*-ten Wiederholung des Experimentes. Das starke Gesetz der großen Zahlen besagt dann, dass es eine Teilmenge E von $\Omega^{\mathbb{N}}$ gibt, die bezüglich des Produktmaßes $\bigotimes_{i \in \mathbb{N}} p_i$ die Wahrscheinlichkeit 1 hat, so dass

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) = \mathcal{E}(X) \quad \text{für alle } \omega \in E$$

oder anders ausgedrückt: Die Menge der $\omega \in \Omega^{\mathbb{N}}$, für die $\frac{1}{n} \sum_{i=1}^{n} X_{i}(\omega)$ nicht gegen $\mathcal{E}(X)$ konvergiert, hat die Wahrscheinlichkeit 0. Deshalb spricht man auch von der Konvergenz fast überall.

ANHANG B. MATHEMATISCHE ERGÄNZUNG
Literaturverzeichnis

- [1] Canny, A computational approach to edge detection, IEEE PAMI 8, 1996, 679-698
- [2] Flannery e.a., Numerical Recipes in C.
- S. Geman und D. Geman, Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, IEEE PAMI 6, 1984, 721-741
- [4] S. Graulund, IEEE Trans. Comp. Volume C-21, 1972, 195-201
- [5] B. Jähne, Digitale Bildverarbeitung, Springer
- [6] H.Niemann, Pattern Analysis and Understanding, Springer 1990
- P. Perona und J. Malik, Scale Space and Edge Detection Using Anisotropic Diffusion, IEEE PAMI 12, 1990, 318-327
- [8] Wallace und Wintz, An efficient three-dimensional aircraft recognition algorithm using normalized Fourier descriptors, Comp. Graphics and Image Processing 13, 1972, 99-126
- [9] J. Weickert und C. Schnörr, PDE-Based Preprocessing of Medical Images, KI 3, 2000, 5-10
- [10] F. M. Wahl, Digitale Bildverarbeitung, Springer 1984
- [11] J. Weickert, Anisotropic Diffusion in Image Processing, Teubner, 1998
- [12] Zahn und Roshies, IEEE Trans. Comp. Volume C-21, 1972, 269-281
- [13] V. Aurich, E. Mühlhaus, S. Grundmann, Kantenerhaltende Glättung von Volumendaten bei sehr geringem Signal-Rausch-Verhältnis, Zweiter Aachener Workshop über Bildverarbeitung in der Medizin 1998.
- J. Weule, V. Aurich, Non-Linear Gaussian Filters Performing Edge Preserving, Proceed. 17. DAGM-Symposium, Bielefeld, Springer 1995, 538-545.
- [15] J. Weule, Iteration nichtlinearer Gaußfilter in der Bildverarbeitung Dissertation, Uni Düsseldorf, 1994.
- [16] C. Tomasi, R. Manduchi, *Bilateral filtering for gray and color images*, Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India.