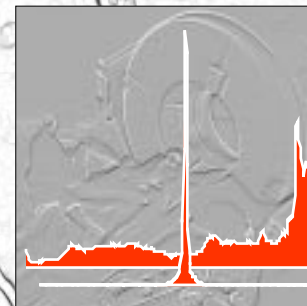
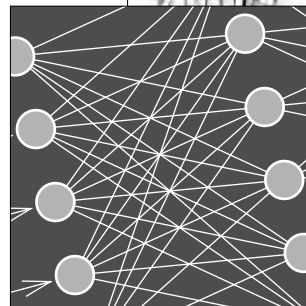
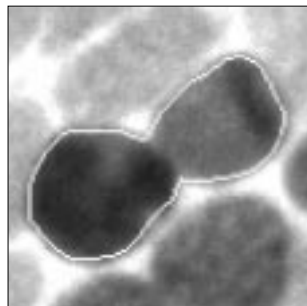
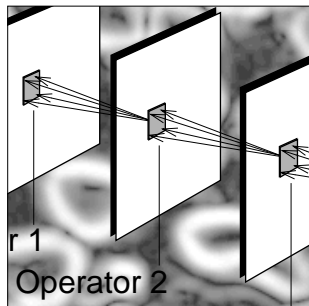


# COMPUTER- VISUALISTIK



## Teil B Bildanalyse



Textbuch zur Vorlesung  
„Digitale Bildverarbeitung II — Bildanalyse“  
V2/Ü1 — Sommersemester

Prof. Dr.-Ing. Dietrich Meyer-Ebrecht  
Lehrstuhl für Messtechnik und Bildverarbeitung  
RWTH Aachen  
Walter-Schottky-Haus (Geb. 24, Erweiterungsgelände Melaten)

Sekretariat: 80.27860 — LfM@RWTH-Aachen.De

Internet: *www.lfm.rwth-aachen.de*  
(aktuelle Informationen zu den Lehrveranstaltungen, Prüfungen etc.)

Version V 2.5 — Oktober 2003

Herausgegeben vom  
Lehrstuhl für Messtechnik und Bildverarbeitung der RWTH Aachen

Alle Urheberrechte liegen beim Autor

Verwendung der Inhalte auch auszugsweise nur mit ausdrücklicher  
schriftlicher Genehmigung des Autors

# Inhalt

<b>Begriffe und Symbole</b>	a
Begriffe zu Teil A	a
Begriffe zu Teil B	b
Formelzeichen und Operatoren	c
<b>Einführung</b>	1
Visualistik – ein neuer Begriff für ein uraltes Thema	1
“Ein Bild sagt mehr als tausend Worte”	1
Computer-Visualistik	3
Ziele und Anwendungsgebiete	5
<b>A Der Abbildungsprozess</b>	11
Das Bild: Verschlüsselung der Realität	11
<b>1 Abbildungsmodelle</b>	15
1.1 Die ‘Abtastung’ des Objektraumes	15
1.1.1 Geometrisches Abbildungsmodell (16)	
1.1.2 Lochkamera (18)	
1.1.3 Zentralprojektion (19)	
1.1.4 Rekonstruktion der Objektposition (20)	
1.2 Unschärfe der Intensitätsmessung	21
1.2.1 Rauschquellen begrenzen die Kontrastauflösung (22)	
1.2.2 Die Relation von Objektgröße und Grenzkontrast (25)	
1.2.3 Quantenrauschen: die ultimative Abbildungsgrenze (26)	
1.3 Unschärfe der Ortszuordnung	29
1.3.1 Die “Point Spread Function” — PSF (30)	
1.3.2 Faltung (32)	

<b>2</b>	<b>Mehrdimensionale Systemtheorie</b>	<b>35</b>
2.1	Aufgaben der Systemtheorie	35
2.2	Die diskrete Fourier–Transformation	36
2.2.1	Ein kleiner Exkurs in die Grundlagen (37)	
2.3	Zweidimensionale diskrete Fourier–Transformation (2D–DFT)	42
2.3.1	Herleitung der 2D–Transformierten aus 1D–Funktionen (43)	
2.3.2	Streifenmuster, harmonische Wellen (44)	
2.3.3	Die 2D–rect–Funktion (45)	
2.3.4	Die zweidimensionale Gauß–Funktion (46)	
2.3.5	Punkt, Punktraster und Gerade (46)	
2.4	OTF— MTF — PTF	48
2.5	Theoreme der 2D–DFT	50
2.5.1	Grundlegende Sätze der 2D–DFT (51)	
2.5.2	Das 2D–Abtasttheorem (52)	
2.5.3	Zentralschnitt–Theorem (53)	
<b>3</b>	<b>Zweidimensionale Bildsensoren</b>	<b>55</b>
3.1	Abbildung — ein Messproblem	56
3.2	Das Vidicon	59
3.2.1	Erzeugung eines Ladungszwischenbildes: physikalische Grundlage (59)	
3.2.2	Die Elektronenstrahl–Abtastung des Ladungsbildes (60)	
3.2.3	Das Signalmodell des Vidicons (61)	
3.3	Der CCD–Matrixsensor	63
3.3.1	Der “Photogate”–Sensor (64)	
3.3.2	Das CCD –Schieberegister (64)	
3.3.3	CCD–Auslesearchitektur (66)	
3.4	Konfokale Laserscanner	68
3.5	Die Gammakamera	71
3.6	Ultraschall–Scanner	72
<b>4</b>	<b>Die 3D–Rekonstruktion aus Projektionen</b>	<b>77</b>
4.1	Dreidimensionale Abbildung — räumliches Sehen	77
4.1.1	Sehen wir dreidimensional? (77)	
4.1.2	Ansätze zur abbildungstechnischen Erschließung der 3. Dimension (80)	

4.2	<b>Computer–Tomographie mit Röntgenstrahlen (CT)</b>	<b>82</b>
4.2.1	Rekonstruktion im Spektralbereich (83)	
4.2.2	“Filtered Backprojection” (85)	
4.2.3	CT — ein abbildendes Meßverfahren (87)	
4.2.4	Anmerkungen zum Rekonstruktionsverfahren (89)	
4.3	<b>Magnetresonanz–Tomographie (MR)</b>	<b>92</b>
4.3.1	Physikalische Grundlage der MR–Bildgebung (93)	
4.3.2	Realisierung einer ortsselektiven Messung (94)	
4.3.3	Das Rekonstruktionsverfahren (95)	
4.3.4	Technische Voraussetzungen (97)	
4.4	<b>Dreidimensionale Abbildung von Isotopenverteilungen</b>	<b>99</b>
4.4.1	SPECT–Bildgebung (99)	
4.4.2	PET–Bildgebung (100)	
<b>5</b>	<b>Restauration und Enhancement</b>	<b>101</b>
5.1	Ziele einer Bild(nach)verarbeitung	101
5.2	Monadische Bildoperationen	103
5.3	<b>Punkt–Operationen</b>	<b>105</b>
5.3.1	Allgemeine Grauwert– bzw. Kontrasttransformation (105)	
5.3.2	Histogramm–Equalisierung (106)	
5.3.3	Adaptive Histogramm–Equalisierung (108)	
5.4	<b>Lokale Operationen</b>	<b>110</b>
5.4.1	FIR–Filter (110)	
5.4.2	Zur Implementierung der diskreten Faltungsoperation (116)	
5.4.3	Rekursive Filter (117)	
5.4.4	Medianfilter (120)	
5.4.5	Optimalfilter (121)	
5.5	<b>Globale Operationen</b>	<b>125</b>
5.5.1	Die DFT (125)	
5.5.2	Filterimplementierung im Spektralbereich vs. Ortsbereich (126)	
5.5.3	Die diskrete Cosinus–Transformation (DCT) (127)	
5.6	<b>Geometrische Operationen</b>	<b>128</b>

<b>B</b>	<b>Bildanalyse</b>	<b>133</b>
	Das Bild — ein komplexes Meßsignal	133
<b>6</b>	<b>Von Pixeln zu Segmenten</b>	<b>141</b>
6.1	Lokale Merkmale	143
6.1.1	Textur (144)	
6.1.2	Statistische Merkmale (146)	
6.1.3	Mikrostruktur-Merkmale (149)	
6.1.4	Spektrale Merkmale (152)	
6.1.5	Wavelets (153)	
6.1.6	Mehrschrittige Merkmals-Extraktion (154)	
6.1.7	Validierung der Merkmals-Extraktion (154)	
6.2	Von Kanten zu Konturen: Kanten-orientierte Segmentierung	155
6.2.1	Die Gestaltgesetze (157)	
6.2.2	Was ist eine Kante? (158)	
6.2.3	Kanten-Detektoren (159)	
6.2.4	Ausdünnung von Kanten zu Konturen (163)	
6.3	Regionen-orientierte Segmentierung	166
6.3.1	Schwellwert-Verfahren (166)	
6.3.2	'region-growing' (168)	
6.3.3	'split-and-merge' (169)	
6.3.4	Signalmodell-basierte Segmentierung (171)	
<b>7</b>	<b>Form und Gestalt: Analyse und Bearbeitung</b>	<b>173</b>
7.1	Von den Pixeln zur Gestalt	174
7.1.1	Die diskrete Geometrie der Pixelwelt (176)	
7.2	Kontinuierliche und diskrete Konturbeschreibungen	178
7.2.1	Kontinuierliche Konturbeschreibungen (179)	
7.2.2	Fourier-Deskriptoren (181)	
7.2.3	Diskrete Konturbeschreibung (182)	
7.2.4	Approximation einer kontinuierlichen Kontur (184)	
7.3	Mit Gestaltkriterien auf Kontursuche	187
7.3.1	Die Hough-Transformation (188)	
7.3.2	Aktive Konturen (190)	
7.3.3	Morphologische Operationen auf binären Bildern (191)	

<b>8</b>	<b>Klassifikation</b>	195
8.1	Der Merkmalsraum	197
8.2	Eine Metrik für die Ähnlichkeit	198
8.3	Das Entscheidungsproblem	199
	8.3.1 Die Bayes-Entscheidungsregel (200)	
	8.3.2 Ein einfaches Beispiel (201)	
	8.3.3 Entscheidungsflächen im mehrdimensionalen Merkmalsraum (201)	
8.4	Überwacht lernende Klassifikatoren	203
	8.4.1 Parametrische Klassifizierer (203)	
	8.4.2 Nichtparametrische Klassifikationsverfahren (204)	
	8.4.3 Neuronale Netze (206)	
8.5	Unüberwacht lernende Klassifikatoren	208
	8.5.1 Die Clusteranalyse (209)	
<b>9</b>	<b>Bilddaten-Kompression</b>	211
9.1	Signal-basierte Kompression: Prädiktionskodierung	213
9.2	Spektrum-basierte Kompression: Transformationskodierung	216
9.3	JPEG — Beispiel eines verbreiteten Standards	217
	9.3.1 'Lossles JPEG' (218)	
	9.3.2 'Lossy JPEG' (218)	
9.4	'Subband coding' — Wavelet-Transformation	219

**Stichworte**

Anhang





# Begriffe und Symbole

## Begriffe zu Teil A

Objekt	:= Körper im Objektraum
Szene	:= Ensemble aus Objekten in einem räumlichen Bezug
Objektraum	:= 3-dimensionaler Raum
Voxel	:= durch Diskretisierung des Raumes definierter kleinster Quader, in dem Objekteigenschaften angebbbar sind (korrespondiert üblicherweise zum <i>Pixel</i> , siehe unten)
Bild	:= über eine Abbildung erzeugte Funktion in einem Bildraum, die bestimmte Eigenschaften eines Objekts oder einer Szene beschreibt
Bildraum	:= Raum, in den ein Objekt oder eine Szene abgebildet wird (oft 2-dimensional, auch 1- oder 3-dimensional, bei dynamischen Abbildungen zeitvarianter Objekte bis 4-dimensional)
Pixel	:= Basisflächenelement im diskreten Bildraum, das sich aus dem Samplingraster ergibt; in 3-dimensionalen Bildräumen wird der Basisquader ebenfalls Voxel genannt
Abbildung	:= Prozess, der Objektvoxel in Bildvoxel bzw. Pixel, d.h lokale Eigenschaften aus dem Objektraum in Bildfunktionswerte transformiert; dieser grundsätzlich mit einem Informationsverlust verbundene Prozeß ist im allgemeinen <i>nicht</i> umkehrbar

## Begriffe zu Teil B

Segment	:= Region räumlich zusammenhängender Pixel, die bezüglich einer bildinhaltsrelevanten Eigenschaft einem Homogenitätskriterium genügen
Kante	:= lokal begrenzte, sich lateral fortsetzende Fluktuation einer bildinhaltsrelevanten Eigenschaft
Kontur	:= Grenze zwischen Segmenten: Konturen müssen <i>geschlossene</i> Linien sein; Beschreibung als 1-Pixel-breite Linie (Pixelkette)
Konnektivität	:= Definition der Pixelnachbarschaft innerhalb geschlossener Flächen bzw. auf linienhafte Pixelketten: auf kartesischen Gittern 4er- oder 8er-Nachbarschaft
Struktur	:= räumliche Zusammenhänge, z.B. Linie, Baum, Netz
Textur	:= kleinräumige, periodisch wiederkehrende Muster — sowohl das Muster selbst als auch seine Periodizität sind durch stochastische Fluktuationen gekennzeichnet
Merkmal	:= aus lokalen Pixelsituationen abgeleitete Eigenschaften (Gradient, Verbundwahrscheinlichkeit, lokale spektrale Maße...)
Operator	:= Algorithmus zur Durchführung einer lokalen Transformation oder zur Bestimmung eines Merkmals — homogener O., inhaltsabhängiger O. (z.B. "Kantenverfolger")
Klassifikation	:= Ordnen von Instanzen in Klassen; Bestimmung der Klassenzahl und -grenzen, "Cluster"
Datenstruktur	:= Datenrepräsentation: Feld, Vektor, String, Baum, Liste...

## Formelzeichen und Operatoren

*Allgemein* Bildkoordinaten sind stets *fini*t und im Fall einer digitaler Darstellung *diskret*

Die Amplitude der Bildfunktion ist im Fall einer digitaler Darstellung *quantisiert*, d.h. die Zahl ihrer Amplitudenstufen ist abzählbar

Bildsensoren haben begrenzte Abmessungen und meist eine abzählbare Anzahl von Sensorelementen

Bildoperationen haben stets einen *finiten Bereich*

$x, y, z$  := Raumkoordinaten, Bildkoordinaten

$\vec{x}$  := Raum- bzw. Bildkoordinaten in Vektorschreibweise

$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \langle x \ y \ z \rangle^T$  := Koordinatenvektor

$\mathbf{M} = \begin{pmatrix} m_{0,0} & \dots \\ \dots & \dots \end{pmatrix}$  := Matrix der Dimension  $i \cdot j$  (z.B. Transformationsmatrix) mit den Koeffizienten  $m_{0 \dots i-1, 0 \dots j-1}$

$\begin{bmatrix} a \\ b \\ c \end{bmatrix}; \begin{bmatrix} a & b & c \end{bmatrix}; \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  := Faltungskerne (z.B. Filter-Operatoren), ein- und zweidimensional mit den Koeffizienten  $a \dots d$

$m, n$  := diskrete Bildkoordinaten

$M, N$  := Bildabmessungen in diskreten Bildkoordinaten

$t$	:= Zeit
$g_1(x, y, z, t)$	:= Beschreibung eines zeitvarianten Objekts oder einer Szene
$g_2(x, y)$	:= 2-dimensionale Bildfunktion
$g_2(m, n)$	:= Bildfunktion im 2-dimensionalen diskreten Bildraum
$h(m, n)$	:= Systemantwort im 2-dimensionalen diskreten Bildraum, "Point Spread Funktion" (PSF)
$f_x, f_y$	:= Ortsfrequenzen von $x$ - bzw. $y$ -Wellen
$k, l$	:= Wellenzahlen in $x$ - und $y$ -Richtung
$G_2(f_x, f_y)$	:= komplexe Fouriertransformierte einer Bildfunktion
$G_2(k, l)$	:= komplexe diskrete Fouriertransformierte einer Bildfunktion
$H(k, l)$	:= komplexe diskrete Fouriertransformierte einer Systemantwort, "Optical Transfer Funktion" (OTF)
$H(k, l)$	:= Betrag der diskreten Fouriertransformierten einer Systemantwort, "Modulation Transfer Funktion" (MTF)
$h_i$	:= Histogramm
$H_i$	:= kumulatives Histogramm
$\hat{g} = E\{g\}, \bar{g}$	:= Erwartung, Mittelwert
$h(\vec{x}) * g(\vec{x})$	:= Faltung
$F\{g\}, F^{-1}\{G\}$	:= Fouriertransformation, inverse Fouriertransformation

# Teil B: Bildanalyse

*Sehen heißt, einige hervorstechende Merkmale von Objekten erfassen* — Rudolf Arnheim

## Das Bild — ein komplexes Mess-Signal

Unsere visuelle Wahrnehmung ist auf das Hier und Jetzt beschränkt. Indem wir visuelle Impressionen als Bild aufzeichnen, das wir transportieren und speichern können, überwinden wir Raum und Zeit. Das Bild kann dabei auch eine spezifische Botschaft enthalten: Es kann Auskunft geben über die Beschaffenheit eines Objekts, über die Konstellation einer Szene, über das Auftreten gesuchter Objekte oder über bestimmte Objekteigenschaften. Um diese Botschaften aus Bildern zu extrahieren, bedarf es einer Entschlüsselung — ein Vorgang, den unsere visuelle Wahrnehmung, wenn sie erst auf das Ziel ausgerichtet ist, so selbstverständlich, so schnell und sicher, so vielgestaltig leistet, dass wir seine Komplexität oft unterschätzen. Immerhin erfordert diese Entschlüsselung eine Verdichtung der Megabits an Daten, die die primäre Bildinformation kodiert darstellen, auf eine Aussage, die sich im Extremfall auf das eine Bit „gut/schlecht“, „ohne Befund/Befund“, „keine Gefahr/Gefahr“ zuspitzt.

Die Leistung der visuellen Wahrnehmung kann erst angemessen gewürdigt werden, wenn wir die Schwierigkeiten erfahren haben, diesen Prozess maschinell nachzuvollziehen. In welchem umfassendem Maße wir Wahrnehmungserfahrung, a-priori-Wissen einbringen, um aus unvollständigen und gestörten Bilddaten ein Objekt oder eine Szene zu rekonstruieren und Eigenschaften, Elemente oder Relationen unscharfen Klassen zuzuordnen, wird uns dabei kaum bewusst. Die daraus erwachsenden Fehleinschätzungen des Entwicklungsaufwandes, um selbst einfach erscheinende Aufgaben zu lösen,

haben in der Vergangenheit zu vielfachen Enttäuschungen über die Leistungsfähigkeit „der“ Computer geführt.

Nun kann man darüber streiten, ob es möglich sein wird oder auch erstrebenswert ist, mit dem Computer unsere visuelle Wahrnehmung zu kopieren. Unbestreitbar kann „maschinelles Sehen“ jedoch in vielen technischen und nicht-technischen Anwendungen unterstützend und ergänzend eingesetzt werden — vor allem, wenn die besonderen Stärken der Computertechnik zur Wirkung kommen können:

- Rechenprozesse sind quantitativer Natur. Sie stützen sich auf quantifizierbare Quellinformation und liefern quantitative Resultate im Gegensatz zu unserer eher qualitativen mentalen Informationsverarbeitung.
- Programmgesteuerte Verarbeitungsprozesse und Funktionsabläufe sind reproduzierbar, unbestechlich und — im Rahmen der gegebenen Fehlersicherheit von Hardware und Software — zuverlässig im Gegensatz zu unseren konditionsabhängigen, durch innere und äußere Situationen beeinflussbaren mentalen Leistungen.
- Maschinelle Informationsprozesse sind in weiten Grenzen skalierbar. Mit entsprechendem Aufwand können sie hinsichtlich Durchsatzmenge und -geschwindigkeit an die Erfordernisse der Anwendung angepasst werden — unbeeinträchtigt durch Ermüdungserscheinungen, nachlassende Aufmerksamkeit in Routinetätigkeiten oder Motivationsverlust in stark taylorisierten Arbeitsorganisationen.

War 'computer vision' seit je ein faszinierendes Forschungsthema — allein schon, weil Bilder als Arbeitsergebnisse per se spektakulär sind —, so blieb es doch mehr als zwei Jahrzehnte ein exotisches Arbeitsgebiet ohne durchschlagende Anwendungserfolge. Schon primitive Lösungsansätze erforderten lange Zeit den Einsatz von Supercomputern oder die Entwicklung problemspezifischer Hardware, um realen Geschwindigkeitsanforderungen zu genügen — beides im allgemeinen mit prohibitiv hohem Aufwand verbunden. Die dramatische Leistungssteigerung der PC-Technologie gegen Ende

des letzten Jahrhunderts gab endlich den entscheidenden Impuls für den Durchbruch der Bildanalyse in Routineanwendungen. Auf dem Markt finden wir inzwischen eine Mannigfaltigkeit preisgünstiger Systeme, die auf Standard-PC-Hardware aufbauen, ausgerüstet mit preiswerten CCD-Kameras und gängigen Multimedia-Komponenten.<sup>1</sup>

Ein Einsatzgebiet, in dem die automatische Bildanalyse mittlerweile etabliert ist, ist die Qualitätssicherung in der industriellen Produktion. An Webmaschinen, Walzen, Extrudern, Schweißautomaten werden Werkstoffe und Produkte schritthaltend mit ihrer kontinuierlichen Be- oder Verarbeitung auf Fertigungsfehler untersucht. Werkstücke aus Metallguss oder Kunststoffspritzguss werden auf verborgene Materialfehler, Form- und Maßabweichungen geprüft — auch hier dem Produktionstakt folgend, um fehlerträchtige Fertigungsprozesse kurzfristig stoppen oder korrigieren zu können.

Ein weiterer industrieller Einsatzschwerpunkt ist die Robotik: 'robot vision' bezeichnet die Ausrüstung von Industrierobotern und Fertigungsautomaten mit Kameras und Bildanalyse-Systemen, die Objekte identifizieren und ihre Lage erkennen können, um die Bewegungsprozesse der Geräte zu steuern.

---

1. So schreibt Pratt im Vorwort seines Standardwerks 'Digital Image Processing', John Wiley & Sons Inc. 1991: „... 1978 I began the preface to the first edition ... with the following statement: 'The field of image processing has grown up considerably during the past decade ... Image processing has found a significant role in scientific, industrial, biomedical, space, and government applications.' Thirteen years later ... I find the quoted statement still valid. The 1980s have been a decade of significant growth and maturity in this field. At the beginning of that decade, many image processing techniques were of academic interest only; their execution was too slow and too costly. Today, thanks to algorithmic and implementation advances, image processing has become a vital cost-effective technology in a host of applications.“ — Zehn Jahre später können wir feststellen, dass dieser Entwicklungsprozess damit keineswegs seinen Abschluss erreicht hatte. Pratt's Statement war auch seinerzeit eher zu optimistisch.

Ein anderes Anwendungsgebiet ist die medizinische Diagnostik. Abbildungsprozesse, die ein umfangreiches und vielgestaltiges Bildmaterial liefern, gehören zum Standardrepertoire einer zeitgemäßen Gesundheitsfürsorge (siehe Abschnitt 3 ab Seite 55 und Abschnitt 4 ab Seite 77). So verwundert es nicht, dass bereits frühzeitig versucht wurde, den Befundungsprozess durch maschinelle Bildanalyse zu unterstützen — bis auf sehr spezielle Anwendungen jedoch bisher ohne den erhofften Durchbruch in die Routinediagnostik. Dies ist sicherlich in der Variationsbreite des „Objekts Mensch“ und seiner individuellen Pathologien begründet. Zweifel sind auch durchaus angebracht, ob der Ersatz des Arztes hier ein erstrebenswertes Ziel ist.

Dennoch eröffnen sich Perspektiven für einen lohnenswerten Einsatz automatisierter Verfahren zum einen in der Präventivmedizin, zum anderen in der Einführung neuer quantitativer Diagnoseverfahren. Untersuchungsmaterial, das mittels Abstrich, Blutentnahme oder Punktion schonend und unter geringem Aufwand entnommen werden kann, kann frühzeitigen Aufschluss über pathologische Prozesse liefern (zum Beispiel degenerierte Zellkerne als Signal für bösartige Tumore). Die visuelle Auswertung entsprechender mikroskopischer Präparate ist jedoch so zeitaufwendig, dass Reihenuntersuchungen auf dieser Basis erst mit Unterstützung einer maschinellen Bildanalyse gesundheitsökonomisch vertretbar werden.

Die maschinelle Bildanalyse spielt darüber hinaus eine entscheidende Rolle bei der Entschlüsselung von Korrelationen zwischen pathologischen Zuständen oder Prozessen und Folgeerscheinungen, die sich in Bildbefunden manifestieren (zum Beispiel eine Veränderung der Gefäßmorphologie infolge systemischer Hypertonie). In vielen Fällen liegen lediglich Hypothesen über derartige Zusammenhänge vor. Eine quantitative Bildanalyse ist in solchen Fällen die Voraussetzung für eine Validierung der zugrundeliegenden Hypothese. Ohne deren weitgehende Automatisierung ist es im allgemeinen aussichtslos, für ein statistisch abgesichertes Ergebnis genügend umfangreiche Studien durchzuführen.



Gerade am letztgenannten Anwendungsszenario wird deutlich, dass Bildinformation im weitesten Sinne als 'Mess-Signal' aufgefasst werden kann — ein hochkomplexes zugegebenerweise! Dem Teilgebiet Bildanalyse der Computer-Visualistik kommt die Rolle zu, die Primärinformation aus den umfangreichen Datensätzen, die der Sensor liefert, zu verdichten und die interessierenden 'Messgrößen' zu entschlüsseln. Dieses Teilgebiet, das vielfach auch als (Bild-)Mustererkennung ('pattern recognition') bezeichnet wird, befasst sich mit den Methoden und Verfahren, die Datenflut der von Bildsensoren und Abbildungssystemen gelieferten Mess-Signale zu den gewünschten Messwerten zu verdichten.

Die Datenverdichtung wird im allgemeinen in Schritten vollzogen. Bei jedem dieser Schritte fließt a-priori-Wissen in den Prozess ein, nach dessen Maßgabe die gegebene Information gewichtet und dem Ziel der Auswertung entsprechend neu kodiert wird. Das Zwischenergebnis ist ein zunehmend kompakterer Datensatz, der jeweils nur noch die für die weitere Verarbeitung relevante Information trägt.

Wenn a-priori-Wissen eingesetzt wird — Modelle der Szenen oder Objekte, Vorgaben für die zu extrahierenden Messgrößen etc. — werden die Lösungsverfahren notwendigerweise aufgabenspezifisch: Jedes neue Problem erwartet eine individuelle Bearbeitung. Die einschlägige Literatur der letzten drei Jahrzehnte zu diesem Thema liefert eine unüberschaubare Fülle von Beispielen.<sup>1</sup> Die Entwicklung war dennoch bisher wenig evolutionär, denn Evolution, das Prinzip der Natur zur Weiterentwicklung der Lebewelt, setzt das Zusammenwirken von Mutation (= innovative Lösungsansätze) und Selektion (= Überleben unter Konkurrenzdruck) voraus. Hoher Aufwand, mangelnde Leistung, fehlende Validierung haben bisher zu oft dazu geführt, dass auch interessante Ansätze im Laborstadium steckengeblieben sind.

---

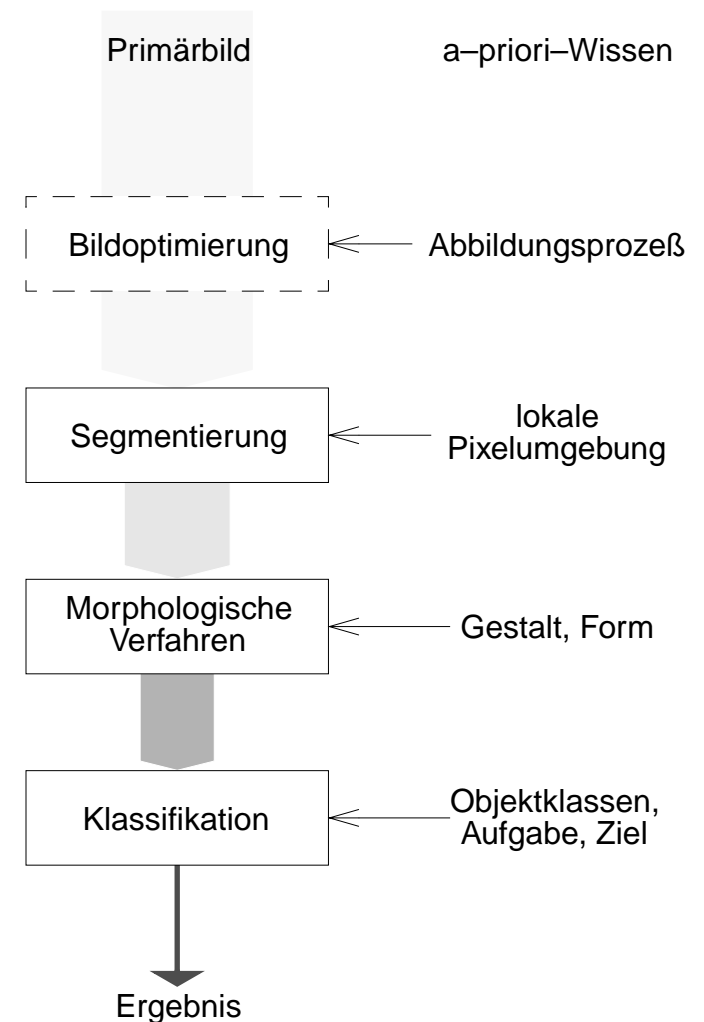
1. Siehe z.B.: USC Annotated Vision Bibliography mit mehr als 20.000 Einträgen

So liegt es in der Natur dieses Arbeitsgebietes, dass eine stringente Systematik bisher nicht formuliert werden konnte. Die Fülle der Methoden und Verfahren kann nur exemplarisch behandelt werden. In den folgenden Abschnitten soll daher eher versucht werden, Denkansätze und Konzepte zu veranschaulichen, die den Verfahren und Lösungswegen zu Grunde liegen.

Unsere Systematik entspricht den aufeinander aufsetzenden Schritten der Datenverdichtung, die durch die Art des jeweils eingesetzten a-priori-Wissens charakterisiert werden können: Nach einer Optimierung des Primärbildes, die sich auf die Parameter des Abbildungsprozesses gründet (ausführlich behandelt in Abschnitt 5 ab Seite 101), fließen zunächst Modelle der lokalen Abbildung relevanter Objekteigenschaften, darauf aufsetzend Beschreibungen der Form und Gestalt, schließlich Aussagen über die Natur der Objekte, über Aufgaben und Ziele der Analyse ein.

Modelle der lokalen Abbildung relevanter Objekteigenschaften dienen als Grundlage, um die vom Abbildungsprozess gelieferte Pixelmatrix zu strukturieren, d.h. alle Pixel so einander zuzuordnen, dass zusammenhängende Bereiche entstehen, in denen sich die abgebildeten Objekte oder Szenen wiederfinden lassen. Innerhalb dieses Prozesses verlieren die Pixel ihre primäre Bedeutung als Träger der Information über die Objektbeschaffenheit in kleinsten abbildbaren Objektraumausschnitten: Sie werden Elemente der Segmente, in die das Bild strukturiert wird, bzw. der Konturen, die die Segmente voneinander abgrenzen. Mit diesem Schritt wird die pixelweise Beschreibung der Szene oder des Objekts aufgegeben zugunsten einer Darstellung aus in sich und miteinander zusammenhängenden Teilflächen (oder -räumen), den Segmenten. Abschnitt 6 „Von Pixeln zu Segmenten“, Seite 141ff, ist den Strategien, Verfahren und Operationen für diesen Prozess gewidmet.

Für die visuelle Wahrnehmung, d.h. für die Einordnung des Gesehenen in Kategorien bekannter Objekte, ist die Erkennbarkeit der Gestalt Voraussetzung — die weitere Interpretation der Segmente und Konturen sowie ihrer



Bezüge zu Szenen und Objekten stützt sich entsprechend auf Beschreibungen der Gestalt. Mit diesem Schritt wird die pixelbasierte Darstellung endgültig verlassen. An ihrer Stelle treten geometrische, parametrische oder analytische Beschreibungen, die als Nebeneffekt zu einer weiteren Datenverdichtung führen. Methoden der Darstellung, Analyse und Verarbeitung von Gestalt-Information sind Inhalt des Abschnitts 7 „Form und Gestalt: Analyse und Bearbeitung“.

Alle Schritte der Datenverdichtung erfordern ein Einordnen, d.h. ein Auftrennen eines quasi-kontinuierlichen Zustandsraumes in eine abzählbare Zahl von Klassen. Dies erfordert eine Entscheidung über die Zuordnung eines Zustandes (Pixel-Funktionswert, Nachbarschaftsbeziehung, Formparameter etc.) zu einer der vorgegebenen Klassen — ein Prozess, der besonders problematisch ist, wenn der Zustandsraum und/oder die Klassengrenzen nur unscharf beschrieben werden können. Aufwendige Klassifikationsverfahren werden vor allem benötigt, wenn, aufsetzend auf die vorhergehenden Schritte, Objekte oder Szenensituationen eingeordnet oder identifiziert werden sollen. Im Abschnitt 8 „Klassifikation“ werden gängige Methoden behandelt.

Die Reduktion der Information eines Bildes auf kompakt darstellbare Messgrößen oder qualitative Aussagen, ist das Wesen der Bildanalyse. Die dabei zugrunde gelegten Betrachtungsweisen und eingesetzten Verarbeitungsprinzipien sind darüberhinaus aber auch hilfreich für die Konzeption von Verfahren, die primär der Kompression von Bilddaten dienen: Wann immer Bilder übertragen oder gespeichert werden sollen, motivieren die üblicherweise sehr umfangreichen Primärdatensätze den Einsatz von Verfahren, mit denen eine kompaktere Kodierung der in den Bildern enthaltenen Information ohne (oder mit tolerabler) Verfälschung erreicht wird. Im letzten Abschnitt 9 „Bilddaten-Kompression“ wird eine Einführung in die Grundlagen und Verfahren zu diesem Thema gegeben.



## 6 Von Pixeln zu Segmenten

In einer große Klasse von Anwendungsfällen ist es Ziel oder notwendiger Zwischenschritt der Analyse, die Gestalt der abgebildeten Objekte in eine grafische Repräsentation zu überführen, d.h. die im Bild identifizierten Objekte allein durch ihre Konturen zu beschreiben — beispielsweise die Zellenquerschnitte in Bild 6—1. Es muss dazu vorausgesetzt werden, dass die abgebildete Szenen aus kompakten Objekten bestehen, die voneinander und gegen leere Umgebung durch einhüllende, d.h. geschlossene Oberflächen abgegrenzt sind. Objekte können kleinere Objekte einschließen oder ihrerseits aus voneinander abgrenzbaren Partitionen bestehen.

Wir gehen nun davon aus, dass ein Abbildungsverfahren angewandt wird, das die durch Materialeigenschaft, Zustand, Dichte, Mikrostruktur etc. definierten Raumpartitionen als in sich zusammenhängend und voneinander unterscheidbar erkennen lässt. Unsere Hypothese für die Bildanalyse lautet folglich:

Geschlossene Flächen<sup>1</sup>, innerhalb derer das Bildsignal oder aus ihm abgeleitete lokale Maße geringe Variationen aufweisen, können als Objekte oder deren Partitionen identifiziert werden. An den Grenzen dieser Flächenbereiche weist die Bildsignaleigenschaft, die das jeweils relevante Homogenitätskriterium erfüllt, eine signifikante Fluktuation auf.

1. Wir schränken uns bei den folgenden Betrachtungen auf zweidimensionale Abbildungen ein. Die Betrachtungen und Verfahren können im allgemeinen auf höherdimensionale Signalräume erweitert werden.

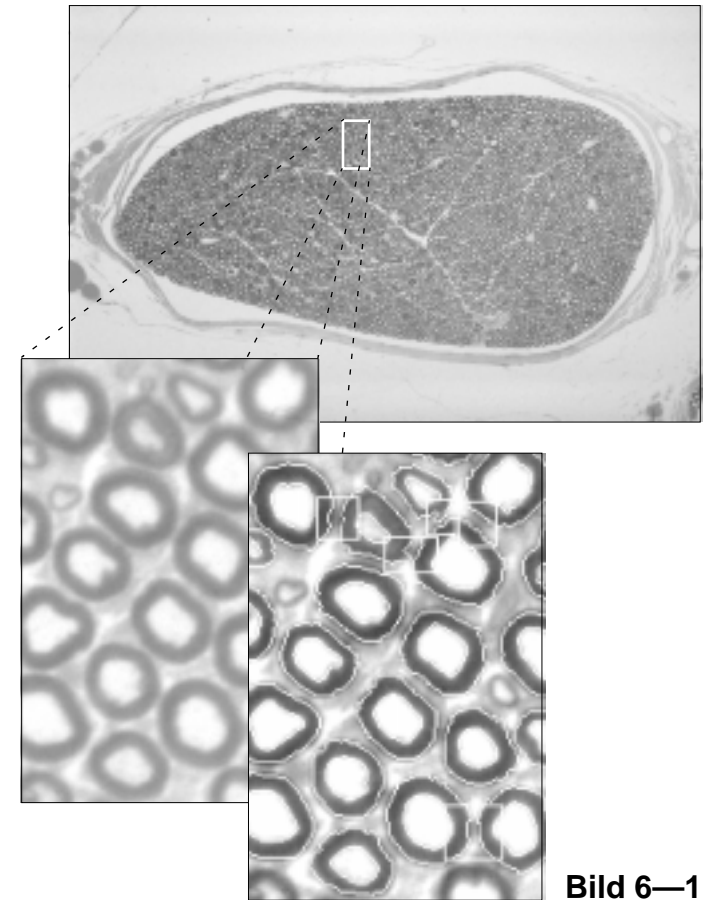


Bild 6—1

Eine zentrale Aufgabe der Bildanalyse ist es, auf der Basis dieser Hypothese eine Ordnung der Pixel des Bildes in zusammenhängende Teilflächen zu schaffen, die in sich (weitgehend) homogen sind. Dieser Vorgang wird *Segmentierung* genannt, die Teilflächen werden als *Segmente* bezeichnet. Die Bildanalyse stützt sich auf dieser Stufe allein auf die Bildsignaleigenschaften. Es kann daher im allgemeinen nicht vorausgesetzt werden, dass eine erfolgreiche Segmentierung bereits eine vollständigere Objekt- oder Szenenbeschreibung liefert. Szenenkonstellationen, Abbildungseigenschaften oder Störeinflüsse können dazu führen, dass ein Objekt in mehrere Segmente zerlegt wird oder dass mehrere Objekte zu einem Segment verschmelzen — Fehler, die erst auf weiteren Stufen der Bildanalyse korrigiert werden können, beispielsweise durch Einfließen von Gestaltinformation.

Den Segmentierungsprozess, wie es das Wort suggeriert, als *Zerlegung* des Bildes zu bezeichnen, leitet unsere Betrachtungen jedoch eher in eine falsche Richtung. Hierin kommt die Perspektive eines Betrachters zum Ausdruck, der den Bildinhalt bereits erfasst hat — eine 'top-down'-Perspektive, die immer wieder zu eklatanten Unterschätzungen der Schwierigkeit von Segmentierungsaufgaben verleitet, da wir uns kaum der komplexen Vorgänge in unserem visuellen Cortex, die Vorstufen und Teile des Wahrnehmungsprozesses sind, bewusst werden.

Aus der 'bottom-up'-Perspektive des Computers ist Segmentierung ein *emergenter* Prozess: Er „erhebt“ die Regellosigkeit der Pixel (die zwar streng an ihren Ort im Bildraaster fixiert sind, aber beliebige individuelle Signalwerte annehmen dürfen) auf eine höhere Ordnung. Die Pixel verlieren ihre Individualität, indem sie Elemente kompakter Flächen werden. Über die jeweilige Zugehörigkeit entscheidet eine Klassifikation: Unmittelbare Pixel-Nachbarschaften oder lokale Pixel-Konstellationen werden flächendeckend daraufhin analysiert, ob sie miteinander verschmolzen werden können.

Die Segmentierung stellt somit den ersten entscheidenden Schritt zur *Datenverdichtung* dar. Anstelle der Wertemenge der individuellen Pixel eines Bildes kann der Bildinhalt nun durch die kompaktere Beschreibung

der Segmente repräsentiert werden: die Konturen der Segmente und das Homogenitätskriterium für die jeweils umfassten Pixel.

Bevor nun Strategien für die Segmentierung behandelt werden, soll geklärt werden, wie die im primären Bild, d.h. in den Pixeln und ihren lokalen Bindungen enthaltene Information möglichst zweckmäßig zur Klassifikation einer Segmentzugehörigkeit beitragen kann.

### 6.1 Lokale Merkmale

Die Segmentierung setzt voraus, dass die Abbildung des Objektes oder der Szene zu einem Bild führt, in dem die Homogenität innerhalb relevanter Teilflächen und die Diskriminanz zwischen den Teilflächen eine genügend sichere Entscheidung über die Zuordnung jedes Pixels zulässt. Im einfachsten Fall kann der Funktionswert selbst (Intensität, Helligkeit, Grauwert des Pixels) die Entscheidungsgrundlage sein. Häufig ist dies jedoch wenig erfolgreich, weil

- systematische oder stochastische Störungen der idealisierten Bildsignalfunktion überlagert sind (Bild 6—2 links: Ausschnitt aus einem Ultraschallbild des linken Ventrikels) oder
- die als Objekt- oder Szenenpartitionen definierbaren Teilflächen nur aufgrund einer einheitlichen Feinstruktur, ihrer sog. *Textur*, als in sich zusammenhängende Flächen erkennbar sind (Bild 6—2 rechts: Dachziegel, Wandziegel, Lattentür, Laub).

In beiden Fällen müssen für die Entscheidung die Relationen des jeweiligen Pixels zu seiner unmittelbaren Umgebung mit herangezogen werden. Nachbarschaftsrelationen, die sich hierfür eignen, werden als *lokale Merkmale* ('local features') bezeichnet. Verfahren zur *Merkmals-Extraktion* ('feature extraction') können im weitesten Sinne als lokale Filter verstanden werden, Bild 6—3: Die Filteroperation, d.h. der Merkmalsoperator produziert ein sog. Merkmalsbild, d.h. ein Datenfeld, dessen Elemente statt der ursprünglichen

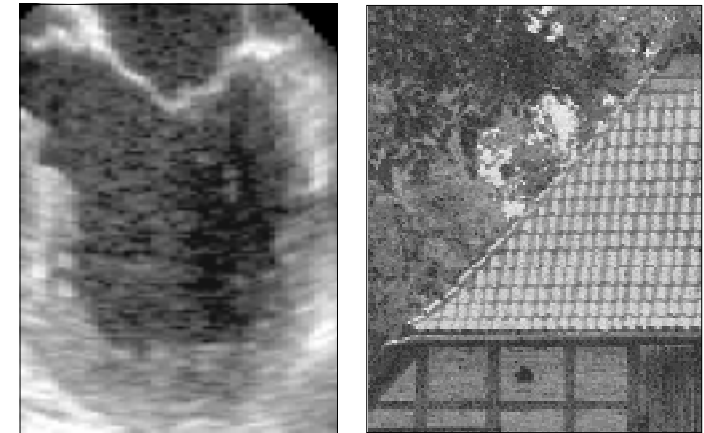


Bild 6—2

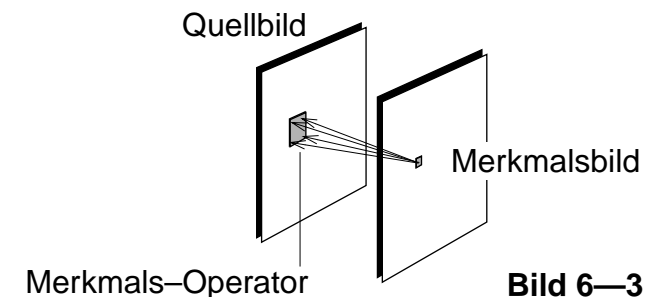


Bild 6—3

Pixelinformation eine aus den Nachbarschaftsrelationen abgeleitete Größe innerhalb einer Region von wenigen Pixeln  $M, N$  um das jeweils betrachtete Pixel herum darstellen.

Im ersten der oben genannten Fälle stören überlagerte lokale Fluktuationen eine durch das Abbildungsverfahren grundsätzlich gegebene Homogenität relevanter Teilflächen. Für die Segmentierung eignen sich in diesem Fall Merkmale, die aus lokalen Mittelwertbildungen entstehen, also mittels signalangepasster Tiefpassfilter gewonnen werden können (bezüglich des Entwurfs und der Implementierung solcher Filter wird auf Abschnitt 5.4 Seite 110 verwiesen).

Im zweiten Fall sind gerade die lokalen Fluktuationen signifikant. Eine entscheidende Maßnahme ist es hier, für die mehr oder weniger offenliegende Systematik dieser lokalen Fluktuationen, der Textur, eine formale Beschreibung zu finden, auf deren Basis ein geeignetes Merkmalsfilter entworfen werden kann. Da Texturen in sehr vielfältiger Weise vorkommen und andererseits aus ihnen sehr wichtige Merkmale für die Segmentierung abgeleitet werden können, soll das Wesen von Texturen und der Umgang mit ihnen zunächst ausführlicher beschrieben werden.

### 6.1.1 Textur

Wenn die Oberfläche eines Objektes aus gemasertem Holz, aus einem genarbten Kunststoff oder aus einem grob gewebten Stoff besteht, ist es sicher nicht Ziel der Segmentierung, die Elemente solcher Oberflächenstrukturen zu differenzieren. Vielmehr kann davon ausgegangen werden, dass innerhalb gestaltrelevanter Flächen eine gewisse Einheitlichkeit derartiger Maserungs-, Narben- oder Gewebestrukturen vorliegt. Ähnliche Situationen können in Luftaufnahmen auftreten, wenn frisch gepflügte Felder sich durch ihr Furchenmuster gegeneinander abgrenzen, oder in Ultraschallbildern, in denen die *Speckle*-Muster (siehe Seite 75) charakteristisch für die Materialeigenschaften sind.

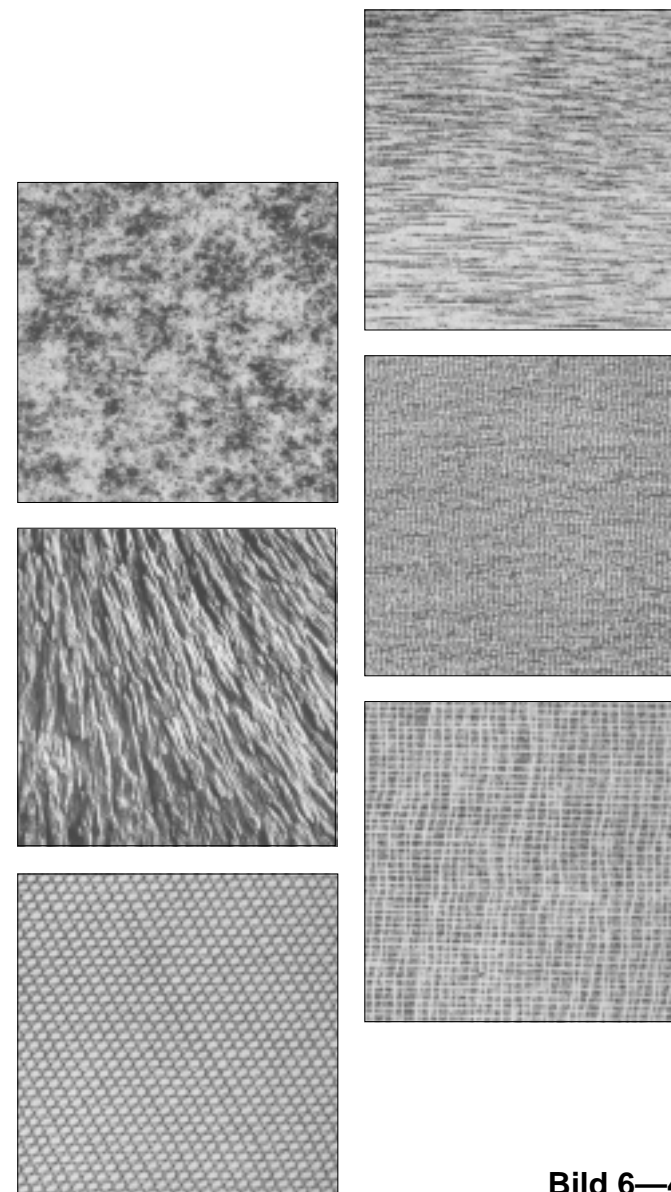


Bild 6—4



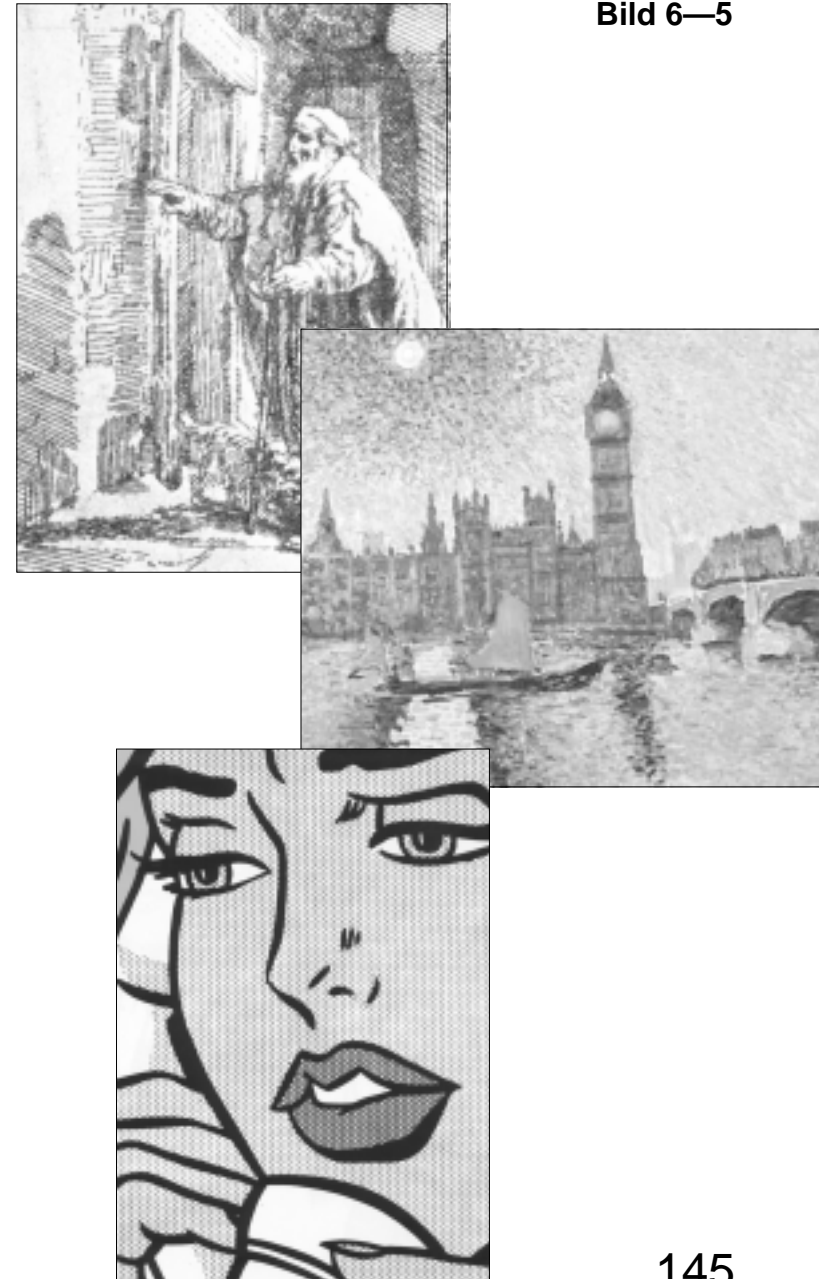
Subsummiert werden alle diese Erscheinungen unter dem Begriff ‘Textur’ (engl. *texture* = Gewebe). Einige Beispiele zeigt Bild 6—4. Die Bandbreite reicht von streng periodischen bis zu rein stochastischen Mustern, von isotropen bis ausgeprägt anisotropen Feinstrukturen. Bei vorwiegend periodischen Texturen sind im allgemeinen die Periodizität und/oder das Mikromuster — periodisch oder quasi-periodisch wiederholte Signalkomponenten — stochastischen Fluktuationen unterworfen. Texturen stochastischer Natur sind oft durch charakteristische, teils schmalbandige Leistungsdichte-Spektren gekennzeichnet.

Wie bedeutungsvoll Texturen für unsere visuelle Wahrnehmung sind, ist unter anderem auch daran zu ersehen, dass sie in vielen unterschiedlichen Kontexten als künstlerisches Ausdrucksmittel genutzt werden — als Darstellungstechnik in Radierungen (Bild 6—5 oben: Rembrandt „Der blinde Tobias“), als Stilelement im Pointillismus der impressionistischen Malerei<sup>1</sup> (Bild 6—5 mitte: André Derain), als visueller Effekt in der Pop-art, um nur einige Beispiele zu nennen (Bild 6—5 unten: Roy Lichtenberg „Oh Jeff...“).

Texturen haben auch in der Kunstphotographie Interesse gefunden. Besonders ist ein Bildband zu erwähnen, den der Fotograf *Brodatz* diesem Thema gewidmet hat<sup>2</sup>: In der Bildverarbeitung hat die in diesem Band zusammengetragene Sammlung von Texturen als Referenz für die Leistungsfähigkeit von Texturanalyse-Algorithmen Bedeutung erlangt.

1. „Betrachtet man nebeneinander stehende Punkte oder Flecken zweier Grundfarben aus einer gewissen Entfernung, so mischen sich diese im Auge des Betrachters zu völlig neuen Farbwerten. Gelbe und blaue Punkte etwa verschmelzen zum Grün, rote und blaue zum Violett. Auch kann eine farbige Fläche eine andere, gleichzeitig wahrgenommene, ohne sie direkt zu reflektieren, ‘aufheizen’. Derlei Erscheinungen des Wahrnehmungsprozesses kannten und nutzten viele Meister instinktiv, um bestimmte malerische Wirkungen hervorzubringen...“ Wolfgang Rainer, Stuttgarter Zeitung vom 19.9.1997
2. Brodatz P: Textures — A Photographic Album for Artists and Designers, Dover Publications, New York, 1966

Bild 6—5



Schließlich zeigt sich bei der Synthese photorealistischer Darstellungen in der Computergrafik, dass Texturen, aufgebracht auf die konstruierten Oberflächen, die dargestellten Szenen deutlich realitätsnäher erscheinen lassen. Auch dies ist ein Indiz dafür, dass Textur eine wichtige Rolle in unserer Wahrnehmung spielt — offensichtlich so zentral, dass die Gestalterkennung durch überlagerte Texturen auch sehr erfolgreich getäuscht werden kann, wie Mimikri und Tarnung in der Natur beweisen. So ist es nicht verwunderlich, dass Texturanalyse-Verfahren zum Grundrepertoire der Mustererkennung gehören.

Texturanalyse-Verfahren haben zum Ziel, lokale Bildbereiche unterschiedlicher Texturen zu differenzieren. Die Konzeption von Texturanalyse-Verfahren setzt auf der Hypothese auf, dass zu Bildinhalten, die wir unter dem Begriff Textur einordnen, sowohl periodische als auch stochastische Signalcomponenten beitragen. Diese Hypothese wird durch Experimente gestützt, in denen mit entsprechenden Signalgemischen erfolgreich Texturbilder synthetisiert wurden. So können auch die Verfahren zur Merkmals-Extraktion in entsprechende Klassen eingeteilt werden, je nachdem, ob die periodischen oder stochastischen Komponenten bezüglich der Charakterisierung vorliegender oder erwarteter Texturen im Vordergrund stehen.

#### 6.1.2 Statistische Merkmale

Wenn bezüglich einer Merkmalsextraktion von statistischen Größen gesprochen wird, sind im allgemeinen Größen gemeint, die ein Histogramm charakterisieren. Histogramme, d.h. diskrete geordnete Darstellungen der Auftrittshäufigkeit, können vom Pixel-Funktionswert (siehe Seite 106) oder von linearen Kombinationen der Funktionswerte von Pixel-Paaren aufgestellt werden. Ihre charakterisierenden Größen streben mit wachsender Zahl der erfassten Pixel (Größe der Bildbereiche oder Anzahl der Bilder gleichartigen Inhalts) Grenzwerten zu: den statistischen Momenten wie Erwartung oder Dispersion (Varianz), siehe auch Seite 22. Werden lediglich die Pixel-

Funktionswerte ausgewertet, spricht man von *Statistiken erster Ordnung* (synonym auch univariate Statistiken). Werden Beziehungen zwischen Pixeln ausgewertet, spricht man von *Statistiken zweiter Ordnung* (auch Verbund- oder bivariate Statistiken).

Für die Gewinnung von Merkmalen zur Texturbeschreibung werden Histogramme  $h_i[g(n',m)]$  über die Pixel innerhalb einer Region von wenigen Pixeln  $M',N'$  um das jeweils betrachtete Pixel herum aufgestellt.  $i = 0 \dots i_{max}$  ist die Menge der diskreten Pixel-Funktionswerte.

Interessante Merkmale aus der Statistik erster Ordnung sind

- Mittelwert  $\bar{g}$ , Gl. (6-1) und
- Varianz  $D$  bzw. Standardabweichung  $\sigma$ , Gl. (6-2).

Mittelwert und Varianz können auch ohne den Zwischenschritt eines Histogramms unmittelbar aus den Bilddaten berechnet werden, Gln. (6-3) und (6-4). Dies empfiehlt sich, wenn  $M' \cdot N'$  nicht wesentlich größer als  $i_{max}$  ist. Die Extraktion des Mittelwertes ist in diesem Fall identisch zu einer rect-Tiefpassfilterung mit einem Faltungskern der Abmessung  $M',N'$ .

Zur Charakterisierung von Texturen liefern Merkmale aus der Statistik zweiter Ordnung meist interessantere Aussagen. Sie beschreiben die Relationen zwischen Pixelpaaren, deren Abstand durch den Abstandsvektor  $\vec{d}$  definiert ist, Gl. (6-5). Die einfachste Verbundstatistik ist das Histogramm über die Differenz  $g_d$  zwischen benachbarten Pixeln, Gl. (6-6).

Für eine Textur-Charakterisierung ist der Mittelwert dieser Statistik, Gl. (6-7), kaum von Bedeutung. Interessanter ist die Standardabweichung  $\sigma_d$ , Gl. (6-8): Sie liefert eine Aussage über den lokalen Feinstrukturgehalt in der durch  $\angle \vec{d}$  gewählten Ausrichtung.<sup>1</sup>

1. Im Fall unmittelbar benachbarter Pixel (Abstandsvektor  $\vec{d} = \langle 1 \ 0 \rangle^T$  bzw.  $\langle 0 \ 1 \rangle^T$ ) kann das Histogramm auch über die Funktionswerte des Gradientenbildes aufgestellt werden, siehe *partieller Gradient* Seite 149.

$$\bar{g} = \frac{1}{M'N'} \sum_{i=0}^{i_{max}} i \cdot h_i \quad (6-1)$$

$$\sigma^2 = D = \frac{1}{M'N'} \sum_{i=0}^{i_{max}} (i - \bar{g})^2 \cdot h_i \quad (6-2)$$

$$\bar{g} = \frac{1}{M'N'} \sum_M \sum_{N'} g(m', n') \quad (6-3)$$

$$\sigma^2 = D = \frac{1}{M'N'} \sum_M \sum_{N'} (g(m', n') - \bar{g})^2 \quad (6-4)$$

$$\vec{d} = \vec{x}_2 - \vec{x}_1 = \begin{Bmatrix} m_d \\ n_d \end{Bmatrix} \quad (6-5)$$

$$g_d = g(\vec{x}) - g(\vec{x} + \vec{d}) \quad (6-6)$$

$$= g(m', n') - g(m' + m_d, n' + n_d)$$

$$\bar{g}_d = \frac{1}{M'N'} \sum_M \sum_{N'} g_d(m', n') \quad (6-7)$$

$$\sigma_d^2 = D = \frac{1}{M'N'} \sum_M \sum_{N'} (g_d(m', n') - \bar{g}_d)^2 \quad (6-8)$$

$$\mathbf{h}_d(i_1, i_2) = \begin{pmatrix} h_{d0,0} & h_{d1,0} & \dots & h_{di_{\max},0} \\ h_{d0,1} & h_{d1,1} & \dots & h_{di_{\max},1} \\ \dots & \dots & \dots & \dots \\ h_{d0,i_{\max}} & h_{d1,i_{\max}} & \dots & h_{di_{\max},i_{\max}} \end{pmatrix} \quad (6-9)$$

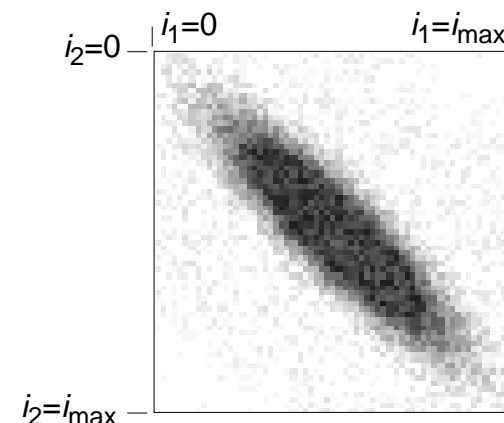
$$d = |\vec{d}| = |\vec{x}_2 - \vec{x}_1|$$

Eine umfassendere Aussage liefert die sogenannte *Grauwertübergangs-Matrix*  $\mathbf{h}_d$  ('cooccurrence matrix'), Gl. (6-9): Der Koeffizient  $h_d(i_1, i_2)$  beschreibt die Auftrittshäufigkeit der Wertekombination  $i_1, i_2$  über alle Pixel-Paare  $(\vec{x}_2, \vec{x}_1)$  mit dem Abstand  $\vec{d}$ . Die Grauwertübergangs-Matrix ist in ihrem Wesen ein zweidimensionales Histogramm über der Wertemenge  $i_{\max}$ . Bild 6-6 (die Koordinatenrichtung ist in der Darstellung der Matrixschreibweise angeglichen, die Grauwerte der „Pixel“ stellen die Auftrittshäufigkeiten  $h_d(i_1, i_2)$  dar).

Im allgemeinen wird die Grauwertübergangs-Matrix rotationsinvariant formuliert, d.h. das Auftreten von Pixel-Paaren  $(i_1, i_2)$  im Abstand  $|d|$  wird über alle Richtungen summiert. Für  $|d|$  wird häufig der einfache Rasterabstand horizontal, vertikal und diagonal gewählt. Um die Matrix nicht zu groß werden zu lassen, wird der Wertebereich im allgemeinen durch Vergrößerung der Quantisierungsschritte reduziert.

Aus der Grauwertübergangs-Matrix können unterschiedliche Merkmale abgeleitet werden, die eine Textur charakterisieren, siehe Gl. (6-10). So liefert die Kovarianz eine Aussage über die Regelmäßigkeit einer Textur. Der Kontrast beschreibt die Grauwertdifferenz und ist damit ein Maß für die Feinheit der Textur. Die Verbund-Entropie ist umso kleiner, je homogener die Textur ist.

Bild 6-6



Energie (6-10)

$$\sum_{i_1}^{i_{\max}} \sum_{i_2}^{i_{\max}} h_d(i_1, i_2)^2$$

Autokorrelation

$$\sum_{i_1}^{i_{\max}} \sum_{i_2}^{i_{\max}} i_1 \cdot i_2 \cdot h_d(i_1, i_2)$$

Kovarianz

$$\sum_{i_1}^{i_{\max}} \sum_{i_2}^{i_{\max}} (i_1 - \overline{g(\vec{x}_1)})(i_2 - \overline{g(\vec{x}_2)}) h_d(i_1, i_2)$$

Kontrast

$$\sum_{i_1}^{i_{\max}} \sum_{i_2}^{i_{\max}} (i_1 - i_2)^2 h_d(i_1, i_2)$$

Verbund-Entropie

$$-\sum_{i_1}^{i_{\max}} \sum_{i_2}^{i_{\max}} h_d(i_1, i_2) \cdot \ln[h_d(i_1, i_2)]$$

6.1.3 Mikrostruktur-Merkmale

Da Texturen grundsätzlich durch lokale Grauwertfluktuationen entstehen, liegt es nahe, zu untersuchen, wie weit letztere für die vorliegenden oder erwarteten Texturen charakteristisch sind. Ein erster Ansatz können *Gradientenbilder* sein: Feingranulare Texturen werden höhere Gradientenamplituden aufweisen als grobgranulare Texturen; bei anisotropen Texturen werden die Gradientenamplituden Vorzugsrichtungen aufweisen.

Im Gegensatz zu eindimensionalen Funktionen ist nun aber der Gradient einer skalaren mehrdimensionalen Funktion ein *Vektor*:  $grad(g)$  oder  $\nabla g$  (*Nabla*-Operator), der Betrag und Richtung der größten Funktionszunahme am betrachteten Ort angibt. Zur Bestimmung des Gradienten einer zweidimensionalen Ortsfunktion werden, unabhängig davon, ob nur der Betrag des Gradienten oder auch die Richtung bestimmt werden sollen, die partiellen Differentiale in den zwei Ortskoordinatenrichtungen benötigt, Gl. (6-11). Ihre diskrete Realisierung führt zu Differenz-Operatoren.

Die einfachste Implementierung eines diskreten partiellen Gradienten-Operators ist die Differenzbildung benachbarter Pixel, Gl. (6-12). Sie kann als Faltung mit den Faltungskernen  $[-1 \ 1]$  oder  $[-1 \ 1]^T$  verstanden werden.<sup>1</sup> Die gerade Anzahl von Koeffizienten erfordert eine Verschiebung des differenzierten Bildes um ein halbes Pixel-Inkrement, d.h. ein Pixel des resultierenden Bildes beschreibt den Gradienten zur wahlweise linken oder rechten bzw. oberen oder unteren Pixelposition. Vermeiden lässt sich diese Verschiebung durch symmetrische Gradienten-Operatoren  $[-1 \ 0 \ +1]$  oder  $[-1 \ 0 \ +1]^T$ . Unterschiede in der Wirkung dieser Gradienten-Operatoren lassen sich aus ihren Übertragungsfunktionen im Spektralbereich abschätzen: Die MTF des symmetrischen Operators lässt auf eine Unterbewertung von Strukturen mit Wellenzahlen nahe der Grenzwellenzahl schließen, Bild 6-7.

1. Darauf zu achten ist, dass Faltungskerne [...] nicht mit Vektoren  $\langle \dots \rangle$  bzw.  $\{ \dots \}$  oder Matrizen (...) verwechselt werden.

$$grad[g(x,y)] = \nabla g(x,y) = \begin{Bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{Bmatrix}$$

$$|\nabla g(x,y)| = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2} \quad (6-11)$$

$$\angle g(x,y) = atan\left(\frac{\partial g}{\partial y} / \frac{\partial g}{\partial x}\right)$$

$$\frac{\partial g}{\partial x} \approx g * [-1 \ 1] \quad \frac{\partial g}{\partial x} \approx g * [-1 \ 0 \ 1]$$

$$\frac{\partial g}{\partial y} \approx g * \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \frac{\partial g}{\partial y} \approx g * \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad (6-12)$$

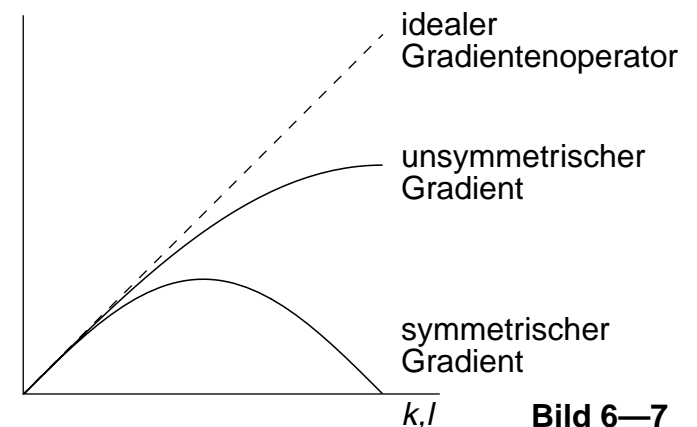


Bild 6-7

Ein partieller Gradient 1. Ordnung, der mit den oben beschriebenen Operatoren in diskreter Form erzielt wird, ist grundsätzlich anisotrop, d.h. richtungsabhängig. Häufig werden jedoch isotrope Gradientenoperatoren gefordert. Für die Konstruktion solcher Operatoren ist das Kriterium die Invarianz gegenüber einer Rotation der Bildfunktion, die durch eine Koordinaten-Transformation nach Gl. (6-13) beschrieben werden kann.

$$\begin{aligned}x' &= x \cos \varphi - y \sin \varphi \\y' &= x \sin \varphi + y \cos \varphi\end{aligned}\quad (6-13)$$

$$g_2 = |\nabla g_1| \approx \left| g^* \begin{bmatrix} -1 & 1 \end{bmatrix} \right| + \left| g^* \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right| \quad (6-14)$$

Das Isotropiekriterium wird durch die Betragsfunktion des Gradienten-Vektors erfüllt, Gl. (6-11). Zur Verringerung des Rechenaufwandes wird die Betragbildung oft vereinfacht zur Summe der Beträge der partiellen Differenzen, Gl. (6-14).

Eine weitere Möglichkeit zur isotropen Gradientendarstellung ergibt sich aus der Realisierung des Gradienten 2. Ordnung, dem *Laplace*-Operator  $\Delta g$ , Gl. (6-15). Dass der Laplace-Operator isotrop ist, wird aus der Fourier-Transformierten der zweiten Ableitung einer Ortsfunktion ersichtlich, Gl. (6-16): Der Term  $(k^2 + l^2)$  beschreibt ein rotationsinvariantes Anwachsen der spektralen Gewichtsfunktion mit dem Quadrat der Wellenzahl.

$$\Delta g(x, y) = \nabla \nabla g(x, y) = \frac{\partial^2}{\partial x^2} (g) + \frac{\partial^2}{\partial y^2} (g) \quad (6-15)$$

$$\mathbf{F}\{\Delta g(x, y)\} = (k^2 + l^2) \cdot G(k, l) \quad (6-16)$$

Für den Laplace-Operator lässt sich eine einfache diskrete Realisierung angeben: Die partielle Ableitung 2. Ordnung kann als zweimalige Faltung der Ortsfunktion mit dem partiellen Gradienten-Operator gemäß Gl. (6-12) formuliert werden. Diese Operation ist identisch mit der Anwendung eines Faltungsoperators nach Gl. (6-17), der wiederum symmetrisch ist. Die Addition der partiellen Gradienten 2. Ordnung in den beiden Koordinaten kann als Faltung mit der Superposition dieser beiden Faltungskerne realisiert werden, Gl. (6-18).

$$\begin{aligned}\frac{d^2 g}{dx^2} &\approx g^* \begin{bmatrix} -1 & 1 \end{bmatrix} * \begin{bmatrix} -1 & 1 \end{bmatrix} = g^* \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \\ \frac{d^2 g}{dy^2} &\approx g^* \begin{bmatrix} -1 \\ 1 \end{bmatrix} * \begin{bmatrix} -1 \\ 1 \end{bmatrix} = g^* \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}\end{aligned}\quad (6-17)$$

$$\Delta g = g^* \left( \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right) = g^* \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (6-18)$$

Eine differenziertere Charakterisierung von Texturen gelingt, indem in der gegebenen Textur nach einfachen kleinräumigen Grundmustern gesucht, d.h. eine Kreuzkorrelation mit geeigneten Musterfunktionen durchgeführt wird. Einen sehr universellen Satz primitiver Musterfunktionen, aus denen durch geeignete Kombination differenziertere Musterfunktionen entwickelt werden können, hat *Laws* angegeben, die sogenannten *LAWS-Textur-energie-Masken* ('lattice aperture waveform sets')<sup>1</sup>, Gl. (6-19).

Da die Kreuzkorrelation bis auf einen Vorzeichenwechsel mit der Faltung identisch ist, Gl. (6-20), kann die Texturanalyse mittels Musterfunktion auch als Filteroperation aufgefasst werden, bei der das Bildsignal mit der Musterfunktion gefaltet wird. Aus diesem Grund werden die Musterfunktionen auch als *Texturfilter* bezeichnet.

Durch die Faltung der in Gl. (6-19) angegebenen eindimensionalen Faltungskerne mit sich selbst oder untereinander entstehen differenziertere und weiter ausgreifende eindimensionale Filter. Zweidimensionale Filter können mittels Faltung eines dieses eindimensionalen Faltungskerne mit einem transponierten Faltungskern konstruiert werden. So ergibt  $[I_3] * [e_3]^T$  den auf Seite 159 beschriebenen Sobel-Operator, ein sehr effizientes Kantenfilter.  $[s_3] * [s_3]^T$  dagegen ergibt den Laplace-Operator mit invertiertem Vorzeichen, siehe Gl. (6-18) oben.

*Laws* gibt weiterhin als Näherung für die Standardabweichung  $\sigma$  den gemittelten Absolutwert der Filterkoeffizienten an. Damit wird die Dispersion  $\sigma^2$  (= Varianz) ein Maß für die Energie der durch die Musterfunktion beschriebenen Textur.

Interessant ist im Vorgriff auf den folgenden Abschnitt die Betrachtung der spektralen Übertragungsfunktionen der Filter. So weisen beispielsweise die Filter  $r_5$  und  $w_5$  ausgeprägtes Bandpass- bzw. Hochpass-Verhalten auf, wie ihre MTF  $R_5$  und  $W_5$  in Bild 6-8 zeigen.

$$I_3 = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \quad (\text{level}) \quad (6-19)$$

$$e_3 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \quad (\text{edge})$$

$$s_3 = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix} \quad (\text{spot})$$

$$I_5 = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \end{bmatrix} = I_3 * I_3$$

$$e_5 = \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \end{bmatrix}$$

$$w_5 = \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \end{bmatrix} \quad (\text{wave})$$

$$r_5 = \begin{bmatrix} -1 & 2 & 0 & -2 & 1 \end{bmatrix} \quad (\text{ripple, oscillation})$$

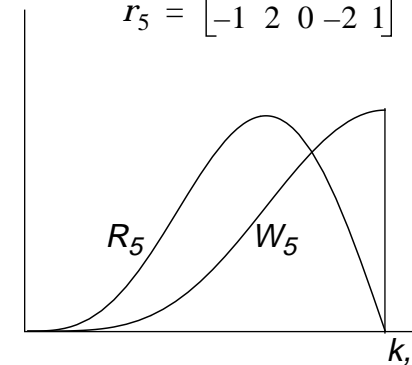


Bild 6-8

1. Laws KI: Textured Image Segmentation, USCIPR Report #940, USC Jan 1980

6.1.4 Spektrale Merkmale

Statistische Bindungen und periodische Strukturen sind — mit jeweils unterschiedlicher Gewichtung — charakterisierend für Texturen. Für die Ansätze zur Merkmals-Extraktion im Ortsbereich lassen sich auf signaltheoretischer Grundlage Entsprechungen im Spektralbereich finden, die zu interessanten Alternativen für die Gewinnung von Merkmalen führen.

Der Ausgangspunkt ist die Feststellung, dass Faltung und Korrelation bis auf die Vorzeichen der Laufindizes übereinstimmen, Gl. (6-20). Ein Vorzeichenwechsel der Laufindizes einer Ortsfunktion führt zur konjugiert Komplexen  $G^*$  der Transformaten dieser Ortsfunktion. So entspricht der Autokorrelation, aus der sich die statistischen Bindungen ablesen lassen, das *Leistungsdichte-Spektrum*  $|G|^2$ .

Eine Merkmals-Extraktion mit den oben beschriebenen Texturenergie-Masken basiert auf einer Kreuzkorrelation des Bildsignals  $g$  mit Grundmustern  $h$ , die für die erwarteten Texturen charakterisierend sind. Die Kreuzkorrelation findet im Spektralbereich ihre Entsprechung in einer Multiplikation des Signalspektrums  $G$  mit der Transformaten  $H^*$  der jeweiligen Musterfunktion, Gl. (6-21). Dies ergibt die spektrale Leistungsdichte der im jeweiligen Bildausschnitt auftretenden, durch die Musterfunktion beschriebenen Signalkomponente (daraus erklärt sich die Bezeichnung *Texturenergie-Masken*, siehe oben).

In beiden Fällen können Merkmale aus den Spektren kleiner Bildausschnitte extrahiert werden. Die DFT ist in einem Bildausschnitt von z.B. 8·8 Pixeln durchzuführen, der pixelweise über den zu analysierenden Bildbereich zu verschieben ist. Als Merkmale können die gemittelten relativen Amplituden charakteristischer Koeffizientenbereiche dienen. So liefern radiale Ausschnitte Aussagen über die Winkelorientierung einer Textur, konzentrische Ausschnitte rotationsinvariante Aussagen über ihre Periodizität, Bild 6—9.

Faltung

$$h(\mathbf{x}) * g(\mathbf{x}) = \sum_{m'} \sum_{n'} h(m', n') \cdot g(m - m', n - n')$$

Kreuzkorrelation

$$h(\mathbf{x}) \star g(\mathbf{x}) = \sum_{m'} \sum_{n'} h(m', n') \cdot g(m + m', n + n')$$

$$h(\mathbf{x}) \star g(\mathbf{x}) = h(-\mathbf{x}) * g(\mathbf{x}) \tag{6-20}$$

$g(m, n)$	○—●	$G(k, l)$	
$g(-m, -n)$	○—●	$G^*(k, l)$	
$g \star g$	○—●	$ G ^2$	
$h \star g$	○—●	$G \cdot H^*$	(6-21)

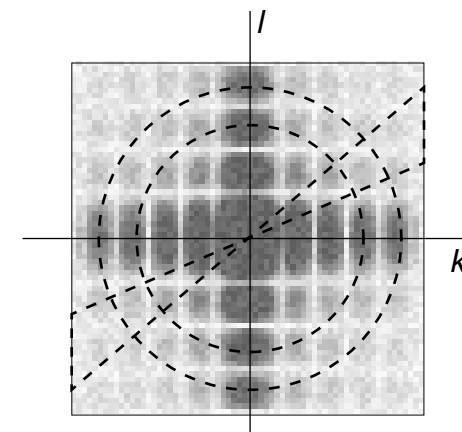


Bild 6—9



Durch die pixelweise Verschiebung des Ausschnitts wird eine unmittelbare Relation der Merkmale zum Ort hergestellt im Gegensatz zur Anwendung der DFT als globale Operation (siehe Abschnitt 5.5 Seite 125), bei der die Lokalisierung der Bildstrukturen in der Phasenfunktion verschlüsselt erscheint.

Die auf diese Weise ermittelten Spektralfunktionen sind allerdings wenig geeignet, um auf die tatsächlich vorliegenden spektralen Komponenten innerhalb der transformierten lokalen Bereiche zu schließen: Je kleiner der Ausschnitt der Bildfunktion ist, der der Transformation unterzogen wird, umso stärker treten Artefakte, die durch die inhärente periodische Wiederholung verursacht werden, hervor. Durch eine Gewichtung des zu transformierenden Bildausschnittes mit einer „weichen“ Fensterfunktion (z.B. *Hamming*-Fenster, Bild 6—10) können diese Artefakte reduziert werden.

### 6.1.5 Wavelets

Die als Bandpassfilter identifizierten Texturenergie-Masken, Seite 151, und die gefensterte Fourier-Transformation innerhalb kleiner lokal-verschiebbarer Bildausschnitte werden durch die *Wavelet-Transformation* ineinander überführt, siehe Abschnitt 9.4 Seite 219. Der Grundgedanke der Wavelet-Transformation ist es, die Unschärfe-Relation der Fourier-Transformation<sup>1</sup> mit einem gleitenden Kompromiss zwischen Lokalität im Ortsbereich und Selektivität im Spektralbereich zu überwinden.

Die Wavelet-Transformation erzeugt eine 'Pyramide' spektral gestaffelter Oktavbänder. Auf den einzelnen Ebenen der Pyramide kann auf Bildauszüge unterschiedlichen Detaillierungsgrades, geordnet in horizontale, vertikale und diagonale Komponenten, zugegriffen werden, Bild 6—11.

1. Lokalität im Ortsbereich führt zu einem ausgedehnten Spektrum, Lokalität im Spektralbereich zu einer ausgedehnten, „unscharfen“ Ortsfunktion. Im Spektrum geht die Zuordnung der spektralen Energie zum Ort verloren.

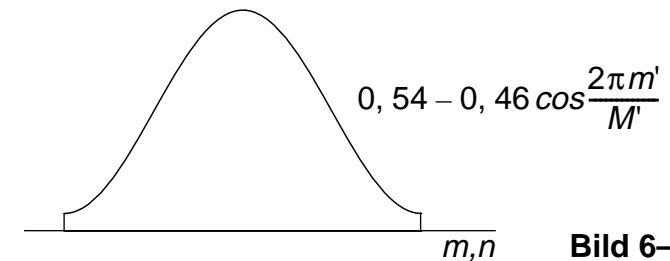


Bild 6—10

Basisbild (Kniegelenk-Röntgenaufnahme)

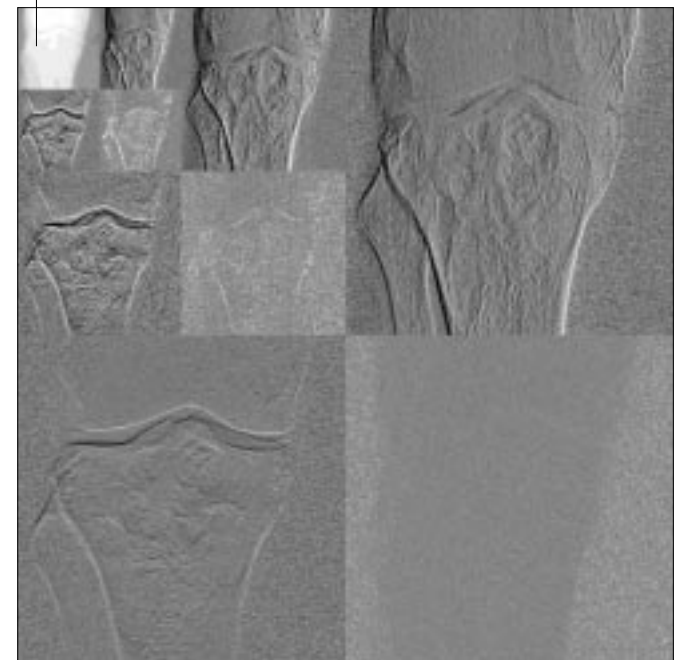
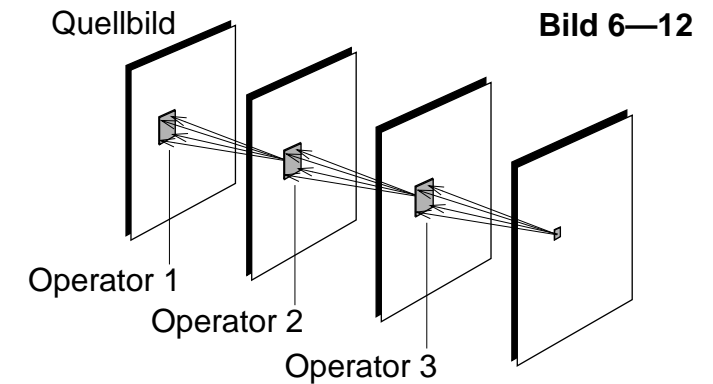


Bild 6—11

## 6.1.6 Mehrschrittige Merkmals-Extraktion

Unterschiedliche Verfahren zur Merkmals-Extraktion können miteinander kombiniert werden. Werden mehrere Verfahren auf dasselbe Quellbild angewandt, dann liefert die Kombination einen Merkmalsvektor, dessen Dimensionalität durch die Anzahl der primären Merkmale gegeben ist. Merkmals-Extraktionsverfahren können auch kaskadiert werden, Bild 6—12. Sinnvoll ist dies, wenn einem sehr kleinräumigen Merkmalsoperator (z.B. einem Laplace-Operator) ein Operator mit ausgedehnterem "Footprint" (z.B. statistischer Operator) nachgeschaltet wird.



## 6.1.7 Validierung der Merkmals-Extraktion

Hinsichtlich ihrer Relevanz können Merkmale nur an vorgegebenem Bildmaterial überprüft werden. Die Kriterien sind

- *Homogenität* innerhalb eines als gleichförmig texturiert wahrgenommenen Bereiches,
- *Diskriminationsfähigkeit* zwischen verschiedenen, als unterschiedlich texturiert wahrgenommenen Bereichen.

Die Grundlage für die Beurteilung können Texturenproben aus problem-spezifischem Bildmaterial sein. Aus Histogrammen von Merkmalsbildern lässt sich mittels der in Abschnitt 8 Seite 195 behandelten Klassifikationsverfahren beurteilen, ob ein gewähltes Merkmal eine genügend sichere Entscheidung erlaubt und ggf. die optimale Entscheidungsschwelle festlegen.

In dem einfachen Modellfall, dass ein oder mehrere einander ähnliche Objekte vor einem Hintergrund zu Segmentieren sind (Beispiel: Mikroskop-Präparat eines Zellastrichs), ist ein *bimodales* Histogramm mit einem deutlichen Minimum anzustreben. Entsprechend sollte der Merkmalsvektor in einem mehrdimensionalen Merkmalsraum gut voneinander trennbare 'Cluster', d.h. lokale Häufungen ausbilden, siehe z.B. Bild 6—26, Seite 165.

## 6.2 Von Kanten zu Konturen: Kanten-orientierte Segmentierung

Das zu segmentierende Bild sei nun — ggf. unter Anwendung geeigneter Merkmalsextraktions-Prozesse — so vorbereitet, dass für die Bildinhalts-relevanten Regionen Homogenitäts- und Diskriminanzkriterien festgelegt werden können. Die Zweigestaltigkeit dieser Kriterien weist bereits darauf hin, dass Strategien zur maschinellen Bestimmung in sich homogener Regionen und deren Grenzen miteinander von zwei Ausgangspunkten aus konzipiert werden können:

- Unter Hinzuziehung der Homogenitätskriterien kann primär versucht werden, Pixelumgebungen zu Regionen zu verschmelzen.
- Gestützt auf die Diskriminanzkriterien können primär Grenzbereiche („Kanten“) zwischen in sich homogenen Regionen bestimmt werden.

Eine entsprechende Klassifikation in *Regionen-orientierte* und *Kanten-orientierte* Segmentierungsverfahren basiert auf der Komplementarität von Region und Rand, Fläche und Umrandung, Raumbereich und Hülle: Ein Segment, das eine geschlossene Fläche darstellt, kann durch die Menge aller zu diesem Segment zählenden Pixel oder alternativ durch seine Kontur, d.h. durch eine geschlossene Linie entlang seiner Berandung bzw. Grenze zu benachbarten Segmenten beschrieben werden. (Die Menge der zu einem dreidimensionalen Segment gehörigen Voxel gegenüber der Beschreibung seiner einhüllenden Oberfläche sind die Entsprechungen im Raum.) Dieser und der folgende Abschnitt werden sich mit den komplementären Klassen der Segmentierungsverfahren befassen.

Angesichts der Selbstverständlichkeit, mit der das visuelle Wahrnehmungssystem Objekte selbst in kompliziertesten Umgebungen (wieder-)erkennt, ist es nicht verwunderlich, dass versucht wird, dem Entwurf von Verfahren für eine maschinelle Mustererkennung vermutete oder experimentell erarbeitete Konzepte der visuellen Wahrnehmung zu Grunde zu legen. Hier spielen Kanten-orientierte Verfahren offenbar eine dominante Rolle. Ihnen wollen wir uns deshalb zunächst zuwenden.

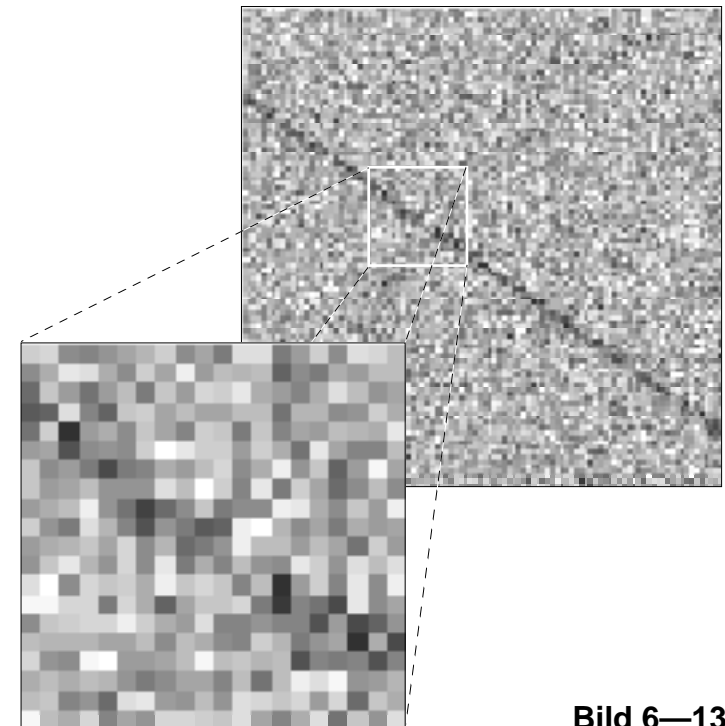


Bild 6—13

Beobachtungen und Experimente führen zu der Hypothese, dass bereits auf einer sehr frühen Verarbeitungsstufe Linienelemente und linienhafte Objektränder aus dem retinalen Primärbild herausgefiltert werden, um diese auf nachgeschalteten Verarbeitungsstufen nach Form- und Umrisskriterien zu analysieren. Linien und Kanten<sup>1</sup>, d.h. Berandungen homogener Flächenbereiche werden selbst bei starken überlagerten Störungen und/oder bei schwachen Kontrasten entdeckt, insbesondere wenn sie von einfacher geometrischer Struktur sind: Obwohl das Muster in Bild 6—13 bei Betrachtung einer kleinräumigen lokalen Umgebung zusammenhanglos erscheint, nehmen wir im größeren Zusammenhang eindeutig eine Linie wahr.<sup>2</sup>

Linien- und Kantenstrukturen sind offensichtlich entscheidend für die Interpretation des Bildinhaltes (zumindest bei der Abbildung natürlichen Szenen). So verfügt unser visueller Wahrnehmungsapparat folgerichtig auch über besondere Fähigkeiten, Linien- und Kantenelemente hervorzuheben. Scheinbare Linien oder Kanten werden sogar dort wahrgenommen, wo sie „ins Bild passen“, d.h. wenn sie der Ergänzung fragmentierter Strukturen dienen, in denen wir bereits ein erwartetes Objekt erkannt haben.

Diese Beobachtung wird durch zahlreiche Experimente mit optischen Täuschungen gestützt. So hat sich auch die Op-art diese Wahrnehmungsphänomene zunutze gemacht: An sich primitive geometrische Strukturen erzeugen bei kaum wahrnehmbaren Deformationen unerwartete, oft sogar überraschende visuelle Erlebnisse, wie z.B. eine scheinbare Perspektive, die eine nichtexistente Tiefe suggeriert (Bild 6—14: Sechseck mit Diagonalen oder Würfel?)

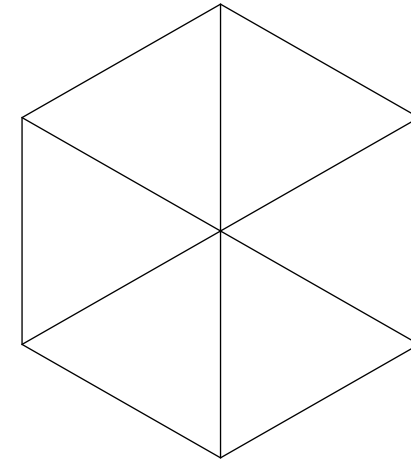


Bild 6—14

- 
1. Zur Definition des Begriffs 'Kante' siehe Abschnitt 6.2.2 Seite 158
  2. Wie wichtig die Konturerkennung für unsere visuelle Perzeption ist, kann bereits aus der Irritation geschlossen werden, die die Betrachtung eines unscharf, verwischt oder verwackelt projizierten Bildes auslöst.

## 6.2.1 Die Gestaltgesetze

Segment-Berandungen, d.h. die meist unscharfen Grenzen zwischen Regionen relativer Homogenität, werden als *Kanten* wahrgenommen und zu *Konturen* (respektive Oberflächen) abstrahiert. Wenn wir diesen Prozess maschinell abbilden wollen, können uns Gesetzmäßigkeiten unserer visuellen Wahrnehmung nützliche Hinweise geben. Diese Gesetzmäßigkeiten wurden bereits zu Beginn des letzten Jahrhunderts von der Wahrnehmungspsychologie untersucht. Die Untersuchungen führten zur Formulierung der sogenannten *Gestaltgesetze*. Einige für unsere Aufgabe interessante Gestaltgesetze, die auch den oben erwähnten optischen Täuschungen zu Grunde liegen, sind im folgenden zusammengestellt.

- *Nähe*  
Elemente, die nahe beieinander liegen, werden als zusammenhängend wahrgenommen:  
Das Punktemuster links in Bild 6—15 scheint aus waagerechten Streifen zu bestehen, das rechte Muster aus senkrechten Streifen.
- *Ähnlichkeit*  
Ähnlich aussehende Elemente werden als eine zusammenhängende Figur wahrgenommen:  
In dem Punktemuster in Bild 6—16 erscheint ein Quadrat, gekennzeichnet durch die vier gleichen, sich von den übrigen Punkten abhebenden Eckpunkte.
- *Geschlossenheit*  
Elemente werden tendenziell als Teile einer geschlossenen Figur interpretiert, auch wenn zwischen ihnen Unterbrechungen auftreten:  
Die aus den Kreisscheiben in Bild 6—17 herausgeschnittenen Segmente werden als überdeckendes Dreieck interpretiert.

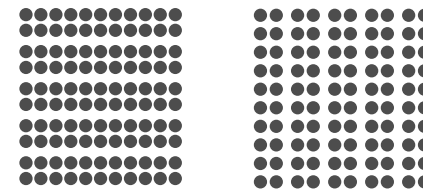


Bild 6—15

Bild 6—16

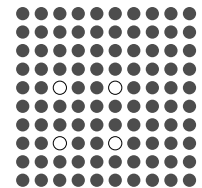
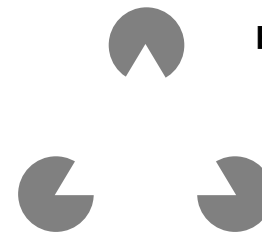


Bild 6—17



- *Gute Fortsetzung*  
Elemente werden als zusammenhängend wahrgenommen, wenn sich zwischen ihnen eine Kontinuität ergibt:  
Die aufeinandergestellten Quadrate in Bild 6—18 erscheinen als eine sich in Form einer Acht überkreuzende, geschlossene Linie.
- *Erfahrung*  
Unvollständige Muster werden ergänzt, wenn dadurch eine bekannte Figur entsteht:  
In Bild 6—19 werden die wenigen Linienzüge zu dem sehr geläufigen Bild des Buchstaben „E“ ergänzt.

Gestaltgesetze werden vermutlich bereits auf einer sehr rezeptornahen Stufe wirksam, um örtlichen Kontext im Bild zu analysieren und daraus Kontur-Information herzuleiten. Sollen Algorithmen entwickelt werden, die Konturen auf der Basis detektierter Kanten finden, muss zunächst geklärt werden, wie sich Kanten signalmäßig darstellen.

### 6.2.2 Was ist eine Kante?

Wenn Segmente nun durch die Homogenität eines geeigneten Merkmals innerhalb ihrer Berandung definiert sind, ist der Rand des Segments bzw. die Grenze zwischen Segmenten gerade durch eine lokale Inhomogenität dieses Merkmals gekennzeichnet. Eine Eigenschaft, die uns hilft, zu einem Rand gehörige Inhomogenitäten von stochastischen Fluktuationen zu diskriminieren, ist ihre *linienhafte Fortsetzung*.

Stellt man sich das zu segmentierende Bild als Geländeprofil vor, bei dem die Geländehöhe den Grauwert oder das ausgewählte Merkmal abbildet, dann stellen sich Segmentränder als Geländestufen dar, Bild 6—20. Die Geländestufe, die keine lokale Erscheinung ist, sondern sich lateral fortsetzt, ist eine geeignete Metapher für den als Bezeichnung für den Segmentrand geläufigen Begriff *Kante*. Wir schließen aus diesem Bild, dass Kantenoperationen so

Bild 6—18

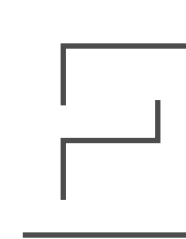
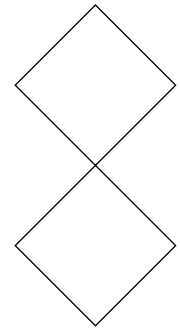
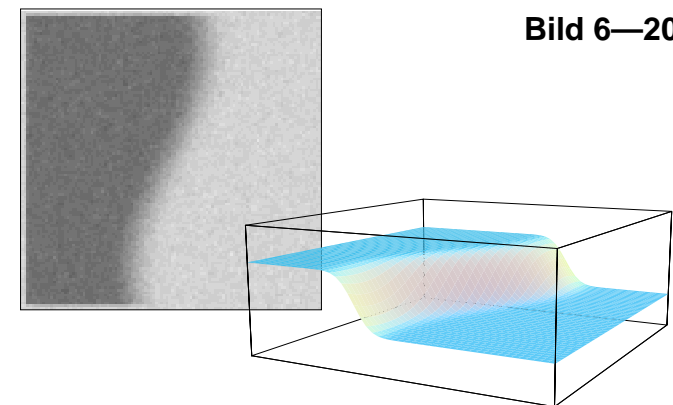


Bild 6—19

Bild 6—20



zu konstruieren sein werden, dass sie verstärkt auf eine laterale Fortsetzung orthogonal zur Richtung des steilsten Gradienten ansprechen.

Eine Kante besitzt nun im allgemeinen auch eine Ausdehnung quer zu ihrer Fortsetzungsrichtung — es sei denn ein stufenförmiger, abrupter Übergang läge vor, was in realen Bildern auf Grund der Abbildungsgesetze (Unschärfe etc.) nicht zu erwarten ist. Die Beschreibung eines Segments durch seine Berandung verlangt jedoch eine linienförmige Definition, d.h., die Kante ist auf eine Pixelkette zu reduzieren, damit sie ein Segment *eindeutig* definiert. Die Pixelkette muss außerdem *geschlossen* sein, damit sie den Rand eines Segmentes *vollständig* beschreibt. Eine auf 1-Pixel-Breite reduzierte und in sich geschlossene Kante bezeichnen wir im folgenden als *Kontur*.

Ein kantenorientierter Segmentierungsprozess besteht folglich aus

1. einer Suche nach Kantenpixeln mittels geeigneter Kantenoperatoren,
2. dem Zusammenschließen von Kanten zu geschlossenen Kantenzügen (dies beinhaltet das Schließen von Zwischenräumen und das Eliminieren scheinbarer Kanten) sowie
3. dem Ausdünnen der Kantenzüge zu Konturen.

Gestaltgesetze geben vor allem für den Schritt 2 nützliche Hinweise.

### 6.2.3 Kanten-Detektoren

Gradienten-Detektoren wie die in Abschnitt 6.1.3 Seite 149 erläuterten sind wenig geeignet zur Kantendetektion, da sie auf jede Art lokaler Fluktuation ansprechen. Eine Bevorzugung von Gradienten, die sich kantenmäßig lateral fortsetzen, kann mit anisotropen Gradientenoperatoren erzielt werden. Ihr klassischer Vertreter ist der *Sobel-Operator*, eine Kombination aus einem eindimensionalen Gradientenoperator und einem dazu orthogonalen eindimensionalen Glättungsoperator (1D-Binomial-Tiefpassoperator), in Gl. (6-22) in seiner kleinsten Version, einem 3·3-Faltungskern dargestellt.

Nach diesem Schema sind auch Operatoren größerer Ausdehnungen zu entwickeln, wenn die erwarteten Kanten eher glatt und/oder verschliffen sind.

Die Faltung mit dem FIR-Filterkern  $h_{Sv}$  nach Gl. (6-22), dessen MTF Bild 6-21 zeigt, kann auch als Kreuzkorrelation verstanden werden, wenn der Faltungskern als Muster eines Kantenelementes gesehen wird. Sie werden daher auch als 'template-based edge detectors' bezeichnet.

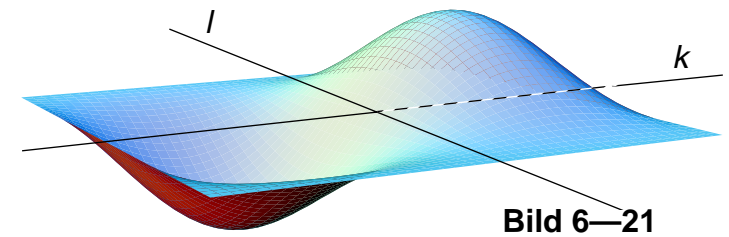
Da mit diesem Filter nur näherungsweise senkrecht ausgerichtete Kanten gefunden werden, wird mindestens ein weiterer Operator  $h_{Sh}$  für horizontale Kanten benötigt. Je ausgedehnter der Filterkern, desto schärfer ist die Richtungs-Selektivität. Deshalb muss ein vollständiger Kanten-Detektor ggf. aus einem größeren Satz richtungsselektiver Operatoren konstruiert werden, z.B. für senkrecht, waagrecht, diagonal-SW/NO und diagonal-NW/SO ausgerichtete Kanten.

Ein solcher Satz von Filtern liefert Aussagen sowohl über die Existenz als auch über die Ausrichtung einer Kante. Als einfaches Merkmal für die Ausprägung einer Kante kann das Maximum der Beträge aller Teilfilter dienen, Gl. (6-23): Die Ausrichtung, d.h. die Richtung des steilsten Anstiegs ist durch den Operator mit der maximalen Amplitude gegeben. Bei Operatorgrößen von wenigen Pixeln macht eine differenziertere Winkelbestimmung wenig Sinn.

Verfeinerte Kanten-Detektoren bauen auf dem Prinzip des Sobel-Operators, d.h. der Kombination richtungsabhängig differenzierender und glättender Filterkomponenten auf. Für eine optimale Trennung zwischen Kanten und stochastischen lokalen Fluktuationen wird der Grundgedanke der Optimalfilter (siehe Abschnitt 5.4.5 Seite 121) herangezogen:

Je eindeutiger sich eine lokale Fluktuation in eine Richtung fortsetzt, je ähnlicher sich ihre Nachbarschaft quer zu dieser Richtung verhält, umso größer ist die Wahrscheinlichkeit, dass hier eine Kante vorliegt.

$$h_{Sv} = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (6-22)$$



$$g_2 = \max(g_1 * h_{Sv}, g_1 * h_{Sh}) \quad (6-23)$$



Mit anderen Worten: Je breiter und glatter eine erwartete Kante ist, desto besser kann sie von stochastischen Fluktuationen unterschieden werden. Zu diesem Zweck wird dem richtungsselektiven Gradientenoperator eine gewichtete Mittelwertbildung in beiden Richtungen überlagert. Die Ausdehnung der Mittelwertbildung respektive die Bandbreite dieses Glättungsfilters wird nach einem Optimalfilteransatz gewählt, der auf die spektralen Eigenschaften der erwarteten Kanten gegenüber den überlagerten Störungen abgestimmt ist.

Regeln für die Dimensionierung eines derartigen Operators hat *Canny*<sup>1</sup> angegeben. Er stellt drei Forderungen auf:

1. Minimierung der Fehlerrate — der Detektor soll nur auf Kanten ansprechen, aber keine Kante übersehen.
2. Optimierung der Lokalisierung — die gefundenen Kanten sollen so nahe wie möglich an der wahren Kontur liegen.
3. Minimierung der Mehrdeutigkeit — wo nur eine Kontur existiert, soll keine Mehrfachkante detektiert werden.

In seinem Kantenmodell geht *Canny* von einer Sprungfunktion mit überlagertem gaußischem Rauschen aus. Die Formulierung der obigen Forderungen laufen auf Signalleistungs-Betrachtungen hinaus, für deren optimale Erfüllung zwar keine analytischen Lösungen vorliegen, jedoch effiziente Näherungen angegeben werden: Das Maximum der spektralen Nutzleistung in Kanten-Längsrichtung liegt sicher bei der Wellenzahl Null. Es wird umso enger um den Ursprung konzentriert sein, je glatter die erwarteten Kanten sind. Daher wird im Spektralbereich auch das SNR-Maximum bei Null liegen. Ein Optimalfilter wird in Kanten-Längsrichtung bei normalverteiltem Rauschen durch eine gaußschen Tiefpass approximiert, dessen Bandbreite von der Kantenstruktur (je weniger glatt, umso breiter) und der Rauschleistung (je höher, umso schmaler) abhängt.

---

1. Canny J: A Computational Approach to Edge Detection, IEEE Trans PAMI Vol C-27 (3) 1986, S 205–213

In Kanten-*Quer*richtung interessiert sicher der Gradient. Das 'Optimalfilter' für diese Richtung besteht folglich aus einer Kaskadierung des für die Kanten-Längsrichtung approximierten Optimalfilters und eines Gradienten-Operators 1. Ordnung.

Diese Kombination aus glättendem und differenzierenden Filtern ist natürlich auch wieder anisotrop. Ein isotroper Kanten-Detektor wird aus zwei orthogonalen Operatoren mit anschließender Bildung des euklidischen Abstands hergestellt, Gl. (6-24). Für die Berechnung einer vollständigen Kanten-Filterung sind dafür jeweils horizontale und vertikale Faltungen mit eindimensionalen Tiefpass-Faltungskernen sowie mit differenzierenden Faltungskernen, die durch Ableitung aus den Tiefpass-Kernen hergeleitet sind, vorzunehmen. Neben dem Betrag des Kantenmerkmals kann auch die Richtung des Gradienten ermittelt werden, Gl. (6-25).  $\sigma$  ist der Optimierungsparameter dieses Kanten-Detektors. Bild 6-22 zeigt die MTF der vertikalen Komponente  $h_n * h'_m$  des *Canny*-Detektors, Bild 6-23 Ergebnisse mit  $\sigma=4$  (links) und  $\sigma=0,9$  (rechts).

*Canny* hat diesen linearen Filteransatz mit nichtlinearen Operatoren erweitert, die Störungen beseitigen und die Kante zur Kontur ausdünnen. Dies soll jedoch Thema des folgenden Abschnitts sein.

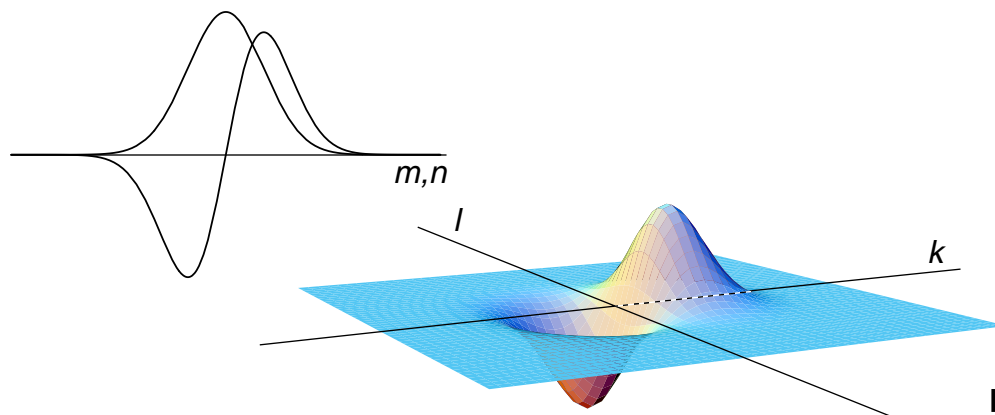


Bild 6-22

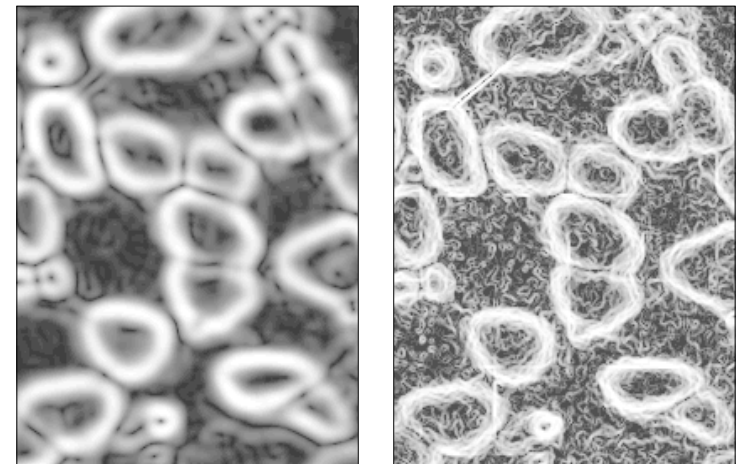
$$|g_{\text{Kante}}| = \sqrt{(g * h_m * h'_n)^2 + (g * h_n * h'_m)^2} \quad (6-24)$$

$$h_m = e^{\frac{-m^2}{2\sigma^2}} \quad h'_m = -\frac{m}{\sigma^2} h_m$$

$$h_n = e^{\frac{-n^2}{2\sigma^2}} \quad h'_n = -\frac{n}{\sigma^2} h_n$$

$$\angle g_{\text{Kante}} = \text{atan} \frac{g * h_n * h'_m}{g * h_m * h'_n} \quad (6-25)$$

Bild 6-23





weniger raffinierte Binarisierung gefunden. Im allgemeinen entstehen abhängig vom Niveau der Entscheidungsschwellen Randunterbrechungen und/oder Kantenartefakte. Jetzt gilt es, Hypothesen für visuelle Wahrnehmungsstrategien — z.B. angelehnt an die auf Seite 157 zusammengestellten Gestaltgesetze — aufzunehmen und in praktikable Algorithmen zu übersetzen, die zur Vervollständigung unterbrochener Konturen und zum Verwerfen der nicht in eine strukturelle Ordnung passenden Konturfragmente führen.

Ein einfacher pragmatischer Ansatz ist der *Hysterese-Operator*, wiederum vorgeschlagen von *Canny* zur Ergänzung seiner 'non-maximum suppression': Mit einem ersten, hoch angesetzten Schwellwert werden Pixel ausgewählt, die mit hoher Wahrscheinlichkeit Konturbeiträge leisten. Die Schwelle wird jetzt reduziert, jedoch werden weiterhin nur noch Pixel in Betracht gezogen, die den vorher ausgewählten Kontur-Pixeln direkt benachbart sind. Dieser Prozess kann iterativ fortgesetzt werden.

Ein raffinierteres Konzept, das auf ein statistisches Verständnis des Bildes aufsetzt, ist die *Kanten-Relaxation*.<sup>1</sup> Das Verfahren beruht auf dem Grundgedanken, dass ein detektiertes Kantenelement umso wahrscheinlicher als wahr erkannt wurde, je stärker Kantenmerkmale in seiner Nachbarschaft so ausgeprägt sind, dass sich eine *beidseitig einfache* Fortsetzung ergibt. In einer Folge von Iterationsschritten werden jeweils alle Elemente eines Kantenmerkmalsfeldes bezüglich ihrer Nachbarschaft geprüft. Je nach lokaler Situation wird das Kantenmerkmal inkrementiert oder dekrementiert.

Die Regeln für diesen Prozess werden im allgemeinen heuristisch aufgestellt, z.B. wie folgt<sup>2</sup>: Alle Pixel eines Gradientenbildes werden zunächst Klassen zugeordnet. Maßgeblich für die Klassifizierung ist, ob sich die Kante in der Normalenrichtung des Gradienten beidseitig, nur einseitig, einfach,

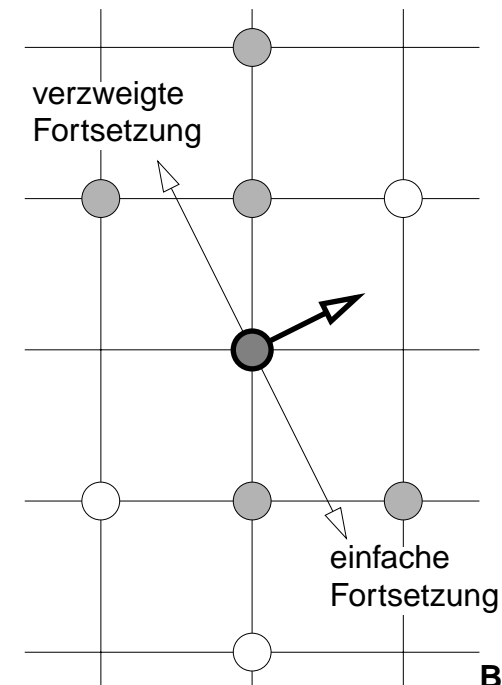


Bild 6—25

1. Unter Relaxation wird allgemein das iterative Auffinden einer kompatiblen Zuordnung von Objekten zu Klassen (hier Kante/nicht-Kante) verstanden.
2. Prager JM: Extracting and Labeling Boundary Segments in Natural Scenes; IEEE PAMI-2 no 1 Jan 1980, 16—27

sich verzweigend oder gar nicht fortsetzt. Dazu werden die Gradienten der in Frage kommenden Nachbarpixel (dünne Kreise in Bild 6—25) auf Überschreitung einer unteren Schwelle des Gradientenbetrags (grau markiert) sowie auf ihre Differenz untereinander geprüft. Bei einer großen Differenz werden die Orte der Pixel mit dem geringeren Betrag, auch wenn deren Betrag die Relevanzschwelle überschreitet, nicht als Verzweigung gewertet. Nach dieser Klassifikation werden die Kanten-Pixel mit beidseitiger einfacher Fortsetzung inkrementiert, Kanten-Pixel mit einseitig verzweigter Fortsetzung dekrementiert und die übrigen Kanten-Pixel unverändert gelassen. Diese Prozedur wird iterativ wiederholt. Der Prozess wird abgebrochen, sobald eine ausgeprägt bimodale Verteilung erreicht ist.

Ein weitergehender Ansatz zur Transformation eines Gradientenbildes in ein Konturbild ist der *Wasserscheiden-Algorithmus* ('watershed algorithm'). Wurden bei den bisher behandelten Algorithmen lediglich direkte Nachbarschaftsbeziehungen genutzt, so geht dieser Ansatz von einer topologischen Interpretation des *Gradientenbildes* aus: Aus einer Kante nach der Definition in Bild 6—20 auf Seite 158 erzeugt eine Gradienten-Operation im Idealfall einen Damm, bei stärkeren überlagerten Fluktuationen, Variationen der Kantensteilheit und anderen Störungen eher eine Art Hügelkette.

Der Wasserscheiden-Algorithmus lässt nun, bildlich gesprochen, die Mulden und Täler, die von diesen Geländeformationen umschlossen werden, voll Wasser laufen, bis die umschließende Dammkrone bzw. Hügelkette an ihrer niedrigsten Stelle überflutet wird. Fortschreitend mit dem weiter steigenden Wasserstand werden immer dort, wo das Wasser weiter zu überfluten droht, auf den Dammkronen bzw. Hügelkämmen Staumauern angelegt und ausgebaut, Bild 6—26. Der Prozess ist abgeschlossen, wenn der höchste Geländepunkt überflutet worden ist. Die Hügelkäme oder Dammkronen sind nun alle durch Staumauern ersetzt, die nunmehr als Konturen gelten können.

Das Umsetzen dieses Konzeptes bringt das Verfahren bereits in die Nähe des 'region growing' (Abschnitt 6.3.2 Seite 168). Allerdings stützt sich der Wasserscheiden-Algorithmus noch ausschließlich auf Kanteninformation.

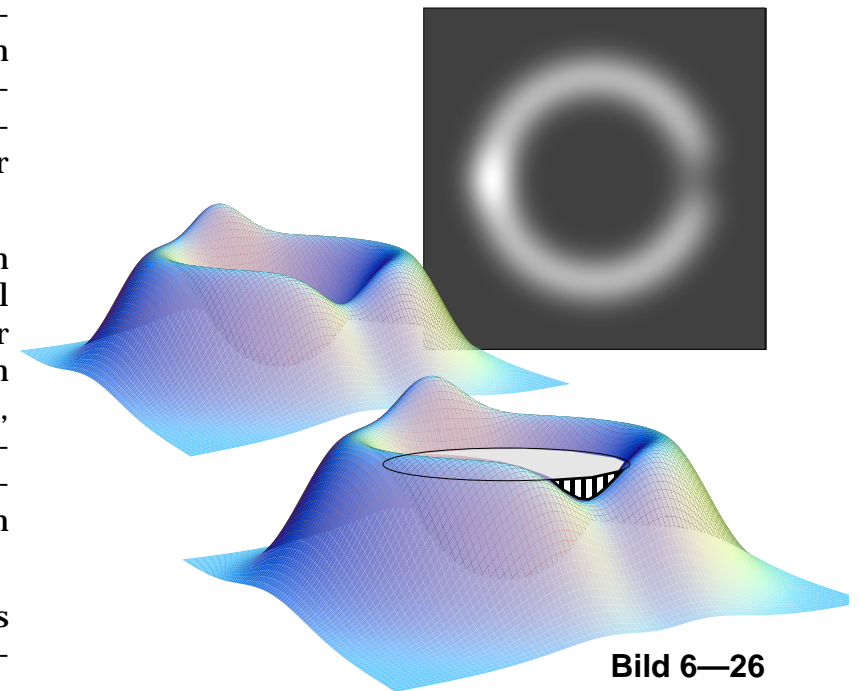


Bild 6—26

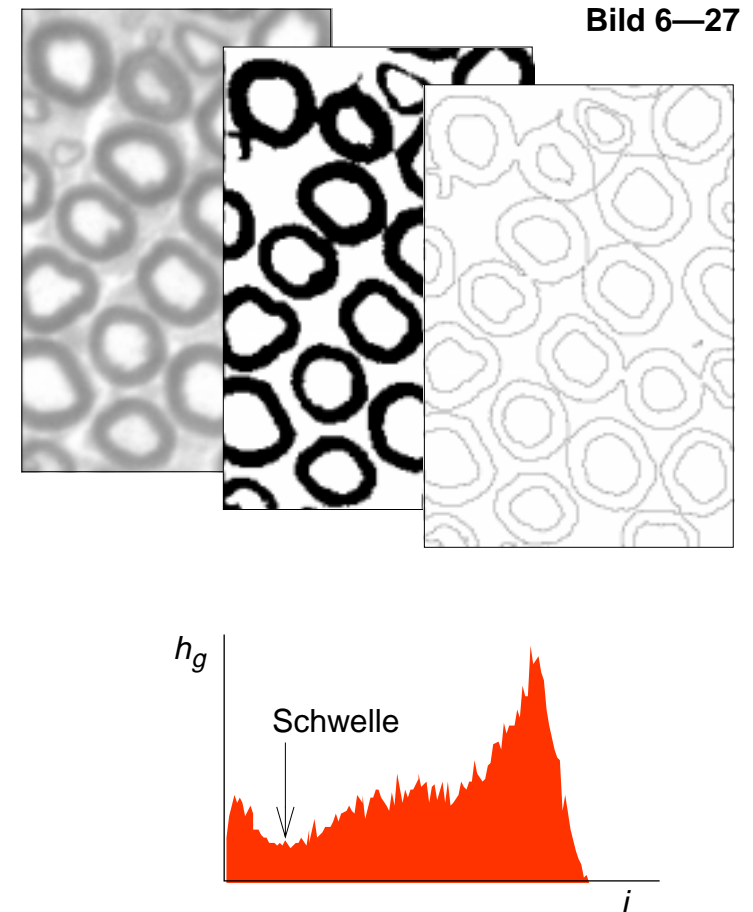
### 6.3 Regionen-orientierte Segmentierung

Anders als der an Hypothesen der visuellen Wahrnehmung angelehnte Kanten-orientierte Ansatz zur Segmentierung folgt der Regionen-orientierte komplementäre Ansatz eher Modellen der Bildsignalentstehung: Wenn mit dem Segmentierungsprozess beabsichtigt ist, Partitionen der abgebildeten Objekte, die sich durch eine Homogenität ausgewählter Merkmale innerhalb kompakter Flächen abzeichnen, zu identifizieren, dann sollte es möglich sein, die zu solchen Flächen gehörenden Pixel aufgrund ihrer Beziehung zu ihrer Nachbarschaft zu Segmenten zu verschmelzen. Die Arbeitshypothese aller Regionen-orientierten Segmentierungsverfahren ist: Pixel, die (1.) benachbart sind und die (2.) einem Homogenitätskriterium genügen, gehören mit hoher Wahrscheinlichkeit zur selben Region. Ihre Menge stellt ein Segment dar.

#### 6.3.1 Schwellwert-Verfahren

In einfachen Situationen kann bereits eine Schwellwertentscheidung zu brauchbaren Ergebnissen führen<sup>1</sup>. Wenn z.B. im Fall isolierter Objekte einheitlicher Natur vor einem näherungsweise uniformen Hintergrund ein ausgeprägt bimodales Histogramm, d.h. ein Histogramm mit einem gut ausgeprägten Minimum des relevanten Merkmals gegeben ist, kann vermutet werden, dass die Maxima die Erwartung dieses Merkmals innerhalb bzw. außerhalb der Objektregionen darstellen. Einen brauchbaren Schwellwert stellt das dazwischen liegende Minimum dar.<sup>2</sup> Bild 6—27 zeigt ein einfaches Beispiel einer solchen Situation (mitte Binärbild, rechts nach Faltung mit Laplace-Operator, siehe Seite 150).

1. Da bei einfachen Schwellwert-Verfahren die Nachbarschaftsbeziehungen nicht betrachtet werden, werden diese Verfahren bisweilen auch zu einer eigenen Klasse der *Punkt-orientierten* Segmentierungsverfahren gezählt.
2. Die Bestimmung optimaler Schwellwerte wird in Abschnitt 8.3.1 Seite 200 behandelt.



Der vom Objekt gegebenen Erwartung sind jedoch Verteilungen infolge stochastischer und/oder systematischer Störungen des Abbildungsprozesses überlagert. So kann eine über dem Bild variierende Grundhelligkeit genauso wie ein hoher Rauschanteil ein an sich ausgeprägt bimodales Histogramm stark einebnen. Alle durch die zu Grunde liegenden Verteilungen gegebenen Fehlentscheidungen beeinträchtigen das Entstehen kompakter Regionen: Falsch-negative führen zu Löchern in Objekt-Regionen, falsch-positive zu Artefakten auf dem Hintergrund. Beide treten oft mit höherer Dichte im Randbereich der Objekte auf. Korrekturmaßnahmen können vor und/oder nach der Schwellwert-Operation eingesetzt werden:

- Vor der Schwellwert-Operation kann eine Glättung erfolgen (Tiefpass-Filter). Stochastische Fluktuationen werden dadurch reduziert. Sind die Objekte vorwiegend glatt berandet, wird die Definition des Randes dadurch wenig beeinflusst. Schwierigkeiten können sich durch ein Zusammenfließen eng benachbarter Objekte ergeben.<sup>1</sup>
- Nach der Schwellwert-Operation können Artefakte mit geringer Ausdehnung durch Median-Filter (Abschnitt 5.4.4 Seite 120) ausgelöscht werden. Umfassendere Bereinigungen können durch morphologische Operationen (Abschnitt 7.3 Seite 187) erzielt werden.

Verbesserungen können auch durch die Hinzuziehung weiterer Merkmale erzielt werden (Farbkanäle eines RGB-Farbbildes, Gradienten etc.) Mittels einer Hauptachsen-Transformation kann versucht werden, ein aussagekräftiges eindimensionales Histogramm aus einem mehrdimensionalen Merkmalsraum zu extrahieren.

Schwellwertverfahren liefern schnell vorläufige Ergebnisse, die leicht über die grundsätzliche Schwierigkeit hinwegtäuschen: Eine 90%ige Segmentierung heisst praktisch, dass die Aufgabe erst zu 10% gelöst ist!

---

1. Durch die Kombination mit Filtern, die über benachbarte Pixel tendenziell Mittelwerte bilden, erhalten auch Schwellwert-Verfahren ansatzweise einen Regionen-orientierten Charakter.

6.3.2 'region-growing'

Ein effektiverer Ansatz, um zusammenhängende Regionen entstehen zu lassen, ist eine Agglomeration benachbarter Pixel, die einem Homogenitätsmaß genügen. Eine bewährte Strategie besteht darin, den Agglomerationsprozess von „Keimpunkten“ ähnlich dem Kristallisationsprozess einer Schmelze quasi simultan fortschreiten zu lassen. Die auf diese Weise definierten Regionen „wachsen“ dabei in das Bild hinein, bis sie auf starke Abweichungen der Homogenität oder auf konkurrierende Nachbar-Wachstumsprozesse treffen.

Zunächst müssen geeignete Keimpunkte ausgewählt werden. Wenn a-priori-Wissen über die erwartete Lage der Objekte nicht vorliegt, kann ein regelmäßiges Gitter von Keimpunkten angelegt werden. Ist das Gitter zu engmaschig gewählt, werden die Agglomerationsprozesse, die von den Keimpunkten ausgehen sollen, häufiger innerhalb derselben Region konkurrieren. Dies führt zu einer *Übersegmentierung*. Ist das Gitter zu weitmaschig gewählt, werden häufiger mehrere Regionen zu einem einzigen Segment zusammengefasst. Es ergibt sich eine *Untersegmentierung*.

An einem Histogramm des Homogenitäts-Merkmals kann überprüft werden, ob der gesamte Wertebereich der Auftrittshäufigkeit entsprechend besetzt ist. Ist dies nicht der Fall, werden weitere Keimpunkte nach einer Zufallsauswahl aus der Menge der Pixel mit den zu dünn besetzten Merkmals-Niveaus bestimmt.

Weiterhin muss ein Homogenitätsmaß festgelegt werden, d.h. bezüglich des relevanten Merkmals muss eine Schwelle gesetzt werden, unterhalb der benachbarte Pixel agglomeriert werden.

Schließlich muss eine Wachstumsstrategie ausgewählt werden, etwa sich spiralförmig um den Keimpunkt entwickelnd oder linear strahlenförmig fortschreitend. Nach Prüfung wird das jeweilige Pixel bei positivem Ergebnis mit dem "label" des Keimpunktes versehen, bei negativem Ergebnis oder wenn auf ein bereits mit einem label versehenes Pixel getroffen wird, wird



das Fortschreiten auf der gewählten Trajektorie abgebrochen. Aus Gründen eines fairen Konkurrierens der von benachbarten Keimpunkten ausgehenden Wachstumsprozesse sollten alle Prozesse alternierend fortschreiten.

Notwendig ist nun eine anschließende Überarbeitung. Benachbarte Segmente werden auf Homogenität geprüft und gegebenenfalls verschmolzen. Mit Stichproben wird überprüft, ob Segmente gegebenenfalls zerlegt werden müssen.

### 6.3.3 'split-and-merge'

Der Ungewissheit, ob Keimpunkte in alle Segmente gesetzt worden sind, die eine vollständige Segmentierung liefern müsste, wird durch die Null-Hypothese ausgewichen, dass jeder Keimpunkt bereits Mittelpunkt einer quadratischen homogenen Region sei. Nach einer Aufteilung des Bildes in ein Schachbrett initialer Regionen (z.B. 32·32 für ein 1024·1024-Bild) wird jede initiale Region auf ihre innere Homogenität geprüft: Ist die Differenz zwischen maximal und minimal auftretendem Wert über alle Pixel einer Region größer als die gesetzte Homogenitätsschwelle, wird diese Region in vier quadratische Sub-Regionen zerlegt. Dieser 'split'-Prozess schreitet gegebenenfalls bis zu Ein-Pixel-Regionen fort, Bild 6—28 oben.

Der 'split'-Prozess kann allerdings noch keine sinnvollen Segmente hervorbringen (es sei denn, die abgebildeten Objekte würden unwahrscheinlicher Weise in eine Schachbrettstruktur passen...). Er stellt jedoch insofern einen interessanten Teilschritt dar, als er die Transformation in eine effizientere Datenstruktur nahelegt: Anstatt jedes einzelne Pixel einer homogenen quadratischen (Sub-)Region mit einem entsprechenden 'label' zu versehen, wird die ganze (Sub-)Region durch einen Knoten in einem Baum repräsentiert. Jede Zerlegung einer (Sub-)Region führt zu einer Vierfach-Verzweigung des entsprechenden Knotens.

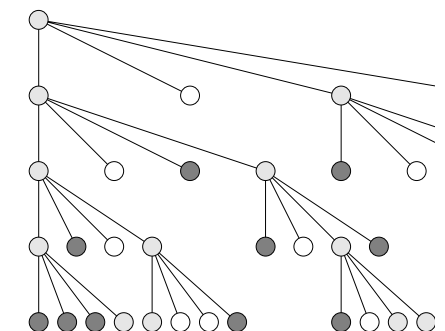
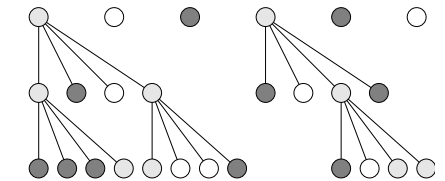
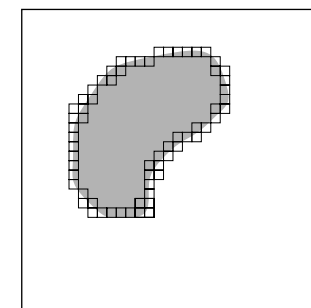
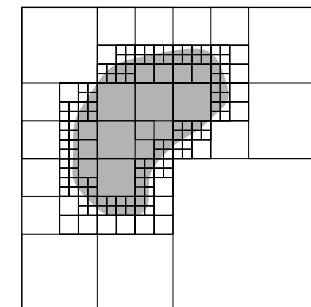
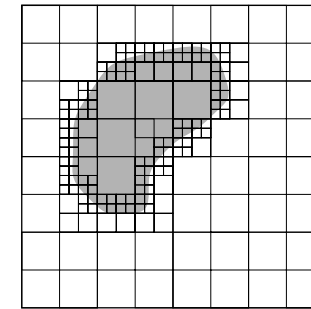


Bild 6—28

Der als Ergebnis des 'split'-Prozesses entstandene 'quad-tree' hat eine Wurzel aus den Knoten der Regionen, von denen der 'split'-Prozess seinen Ausgang genommen hat. Seine Verzweigung führt jeweils zu einer Detaillierung, die eine beste Anpassung der Sub-Regionen an eine Segmentgrenze erlaubt. Die Blätter (alle nicht weiter verzweigten Kanten des Baumgraphen) tragen als Attribut den Mittelwert der als homogen klassifizierten jeweiligen Sub-Region sowie implizit deren Bildkoordinaten. Auf der untersten Ebene finden sich außerdem die Kontur-Pixel, wenn diese Klasse definiert ist. Außer einer komprimierten Darstellung des Bildinhaltes liefert die Baumstruktur einen algorithmisch einfachen Ausgangspunkt für die nun folgende Zusammenfassung von Sub-Regionen zu Segmenten — dem 'merge'-Prozess.

Zunächst wird die initiale willkürliche Aufteilung des Bildes aufgehoben, indem die entstandene 'quad-tree'-Konstellation auf der Wurzelebene zu einem vollständigen 'quad-tree' ergänzt wird, dessen Wurzel das (quadratische) Bild als Ganzes ist. Alle Knoten, deren vier Kanten zu Blättern mit homogenen Attributen führen, werden dabei zu Blättern reduziert — ein erster 'merge'-Schritt, Bild 6—28 mitte.

Der 'quad-tree' kann jetzt in einem weiteren 'merge'-Schritt zu einem Baum umstrukturiert werden, dessen oberste Knotenebene die vollständige Segmentierung repräsentiert. Dazu werden zunächst jeweils alle Blätter eines Knotens auf Erfüllung des Homogenitätsmaßes überprüft und ggf. zu eigenen Knoten zusammengeführt. Hierarchisch aufsteigend werden nun Knoten und/oder Blätter, die aneinander angrenzende Sub-Regionen repräsentieren, ihrerseits auf Erfüllung des Homogenitätsmaßes überprüft und ggf. zu eigenen Knoten zusammengeführt. Dieser Prozess mündet in Knoten, die die Segmente repräsentieren. Sind Kontur-Pixel definiert, können diese durch einen weiteren Knoten repräsentiert werden. Von einem der obersten Knotenebene ausgehend können auf Grund der impliziten Lokalisierung alle dem entsprechenden Segment bzw. die seiner Kontur zugehörige Pixel identifiziert werden, Bild 6—28 unten.

6.3.4 Signalmodell-basierte Segmentierung

Der Segmentierung war eingangs die Hypothese zu Grunde gelegt worden, dass Bildregionen, die für die Interpretation der Objektgestalt oder der Szenenkonstellation relevant sind, homogen und weitgehend zusammenhängend sein müssten. Durch örtliche Helligkeitsschwankungen der Beleuchtungsquelle, durch winkelabhängige Reflexion oder durch einen ungleichförmigen Querschnitt eines durchstrahlten Objektes werden häufig großflächige Signalfluktuation hervorgerufen. Innerhalb der Regionen sind in diesem Falle zwar nicht mehr konstante, aber dennoch 'glatte' Signalverläufe zu erwarten.

Ausgehend von einem solchen Signalmodell kann versucht werden, die vorliegenden Signalverläufe durch zweidimensionale Polynome möglichst niedrigen Grades zu approximieren (in dem konstruierten Beispiel in Bild 6—29 werden Polynome 2. Grades verwendet). Dies gelingt besonders gut, wenn ausgedehnte Bildbereiche von einem uniformen, aber nicht notwendigerweise konstanten Hintergrund ausgefüllt werden. Begonnen wird in diesem Fall mit der Anpassung eines Hintergrundmodells, d.h. es wird eine Polynom-Approximation für das gesamte Bild gesucht. Alle ein Fehlermaß überschreitenden Teilregionen werden herausgeschnitten. Innerhalb dieser Regionen werden wiederum Polynomapproximationen durchgeführt. Der Prozess wird iterativ fortgesetzt, bis keine Fehlerregionen mehr auftreten.<sup>1</sup>

1. Erfolgreich wurde auf diese Weise der Blutgefäßbaum in Bildern des Augenhintergrundes segmentiert (Kaupp A: Ein Verfahren zur Morphologieanalyse des retinalen Gefäßbaumes, Dissertation am Lehrstuhl für Messtechnik und Bildverarbeitung, VDI-Verlag Reihe 17, Nr. 131, 1995, Bildbeispiel siehe Bild 5—18 auf Seite 112)

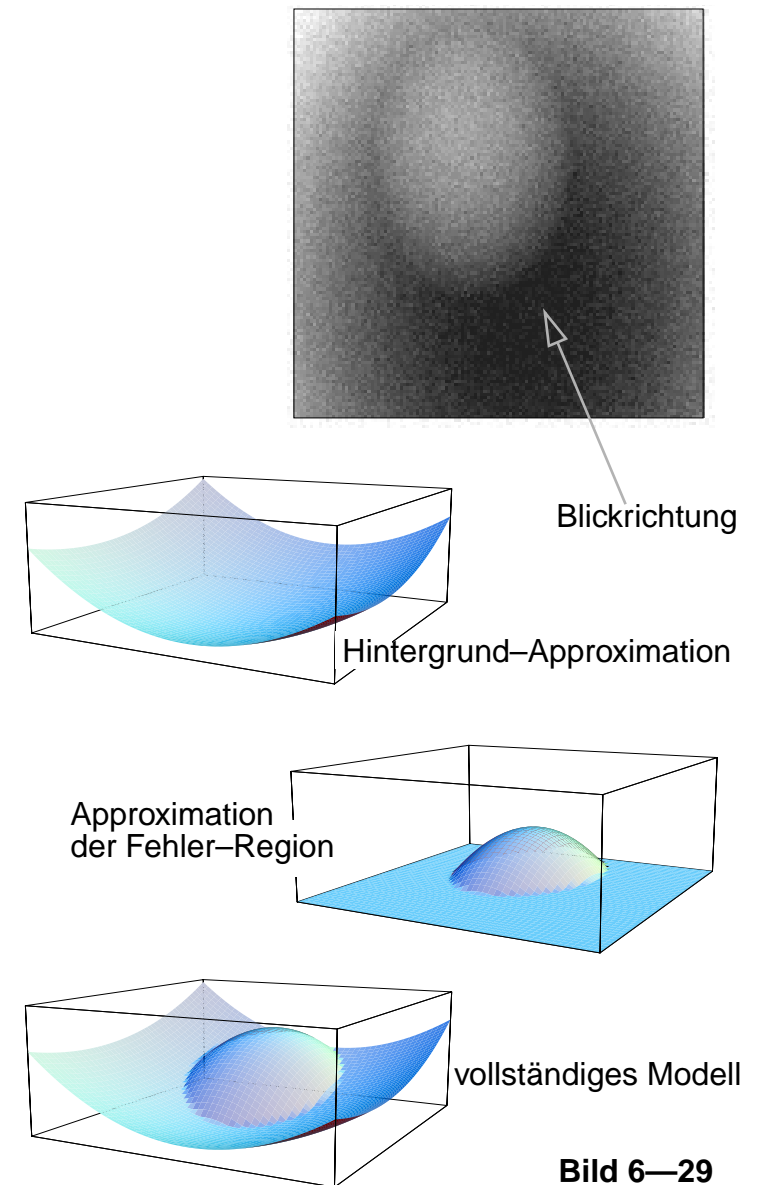


Bild 6—29

*Regionen versus Kanten* — ob ein Regionen- oder ein Kanten-orientierter Ansatz erfolgreicher ist, hängt von der Aufgabenstellung ab. Erfahrungsgemäß neigen Regionen-orientierte Verfahren zur Übersegmentierung, da den Agglomerationsmechanismen die Konkurrenz innewohnt. Kanten-orientierte Verfahren (mit Ausnahme des Wasserscheiden-Algorithmus) tendieren dagegen zur Untersegmentierung, da eine Sicherheit einer vollständigen Detektion aller Konturen nicht gegeben werden kann.

Verbesserungen können erst erzielt werden, wenn a-priori-Information auf der morphologischen Ebene, d.h. Wissen über die erwartete Gestalt der zu segmentierenden Objekte hinzugefügt wird. Eine formale Beschreibung der Gestalt der erwarteten Objekte bietet eine Möglichkeit, die Güte einer Segmentierung zu überprüfen und Optimierungskriterien für ihre Korrektur aufzustellen. Deren Ziel kann es sein, in einem übersegmentierten Bild Regionen zu verschmelzen oder in einem untersegmentierten Bild Konturen zu ergänzen, wenn dadurch jeweils Regionen geschaffen werden, die den vorgegebenen Gestaltkriterien besser genügen. Der Verarbeitung der Bildinformation auf der morphologischen Ebene ist der nächste Abschnitt gewidmet.

## 7 Form und Gestalt: Analyse und Bearbeitung

Unsere visuell erfassbare Welt ist eine Welt der gestalthaften Objekte. Amorphe Erscheinungen — Erscheinungen ohne Gestalt, griechisch  $\mu\omicron\rho\phi\eta$  — können wir visuell nicht wahrnehmen. Der Begriff 'Gestalt' lässt sich andererseits aus der Erfassbarkeit eines Objektes deuten: Gestalt steht für „Figuren oder Fakten, die nur aus der Gesamtheit, d.h. als geschlossenes, sich von einem Grund abhebendes Ganzes zu verstehen sind“.<sup>1</sup> Verborgen ist in dieser Definition die visuelle Dominanz unserer Wahrnehmung: Unmittelbar ertasten oder mittelbar über rückgestreutes Licht visuell erfassen können wir die Form eines räumlichen Objektes nur, wenn es sich als kompaktes Gebilde durch eine einhüllende Oberfläche von seiner Umgebung abgrenzt. Er-fassen, be-greifen: In diesen Begriffen steckt die Bedeutung der taktilen, haptischen und motorischen Raum-Erfahrungen für die Entwicklung unserer visuellen Wahrnehmung. Eng mit 'Gestalt' hängt auch der Begriff 'Form' zusammen. Die Form ist einerseits die formende Hülle (z.B. die Gussform), andererseits kommt in diesem Begriff das Formale, zum Beispiel mit Termini der Geometrie Beschreibbare zum Ausdruck.

Im Zweidimensionalen reduziert sich die Grenzfläche des räumlichen Objektes — seine "Hülle" — zu einer *Grenzlinie*, der Berandung oder Kontur. Im Abschnitt 6.2 ab Seite 155 wurde bereits beschrieben, wie wichtig Gestalt- bzw. Formmerkmale von Objektkonturen für die visuelle Bildinterpretation sind; hingewiesen wurde auf die *Gestaltgesetze* (Abschnitt 6.2.1, Seite 157). Dort wurde bereits die Hypothese aufgestellt, dass systematische geometrische Zusammenhänge wie Linien und Kanten für die Interpretation einer abgebildeten Szene ausschlaggebend sind. Aus unseren alltäglichen Beob-

---

1. Meyers Lexikon

achtungen können wir schließen, dass in natürlichen Szenen stochastische Linien- und Kantenverläufe eher unwahrscheinlich sind. Vermutlich wird, so folgern wir daraus, unser visuelles System bevorzugt nach glatten Fortsetzungen und ausgeprägten lokalen Richtungsänderungen wie Ecken suchen, um im wahrgenommenen Bild eine strukturelle Ordnung zu entdecken, die mit bereits Erfahrenem abgeglichen werden kann.

Es lässt sich nun vermuten, dass wir aus diesen Beobachtungen und Hypothesen Hinweise für Strategien der Bildanalyse gewinnen können, mit denen wir uns von der Ebene der Pixel ablösen und das Bearbeitungsziel auf übergeordnetere Zusammenhänge richten können. Unter diesem Aspekt wird der vorliegende Abschnitt die formale Darstellung der Form bzw. der Gestalt eines Objektes behandeln. Darauf aufbauend werden Ansätze für die algorithmische Umsetzung entsprechender Strategien entwickelt, die zu verschiedenen Klassen von Operationen für die Analyse und Bearbeitung der Gestalt führen.

## 7.1 Von den Pixeln zur Gestalt

Die mentale Leistung unserer visuellen Wahrnehmung stützt sich offensichtlich ganz wesentlich auf vorgelagerte Prozesse, mittels derer strukturelle Ordnungen in die primären Reizmuster gebracht werden. Hier spielen die Gestaltgesetze eine wichtige Rolle, insbesondere das Prägnanzprinzip, auch das *Gesetz der guten Fortsetzung* genannt:

Unter allen Möglichkeiten des Zusammenschlusses von Gliedern zu einem Ganzen wird diejenige bevorzugt, die die einfachste und deutlichste Gestalt darstellt.<sup>1</sup>

- 
1. Wenn andererseits unter dem Begriff Gestalt bereits „das zusammenhängende Ganze“ verstanden wird, ist dieser Satz selbstbezüglich — worin sich einmal mehr die Schwierigkeit der Definition dieses Begriffes offenbart.

Von Pixeln zu Form und Gestalt zu gelangen, ist ein hierarchischer Prozess. Der „Zusammenschluss der Glieder...“ beginnt auf der untersten Ebene der Pixel: Dort werden Pixel zunächst zu geschlossenen Regionen bzw. Konturen verschmolzen — der Prozess der *Segmentierung*, der Thema des vorangehenden Abschnitts 6 „Von Pixeln zu Segmenten“ war.

Die bisher behandelten Strategien blieben der Pixelmatrix als bildhafte Repräsentation des Verarbeitungsergebnisses verhaftet: Das Ergebnis des Segmentierungsprozesses war eine Etikettierung der Pixel ('label'), um sie als Element einer Kontur oder eines Segments zu kennzeichnen. Unser Ziel muss es nun sein, uns von der Pixelmatrix zu lösen und aus den segmentierten Bildbeschreibungen zu erarbeiten, aus denen Form- und Gestaltmerkmale extrahiert werden können. Dieser Abschnitt widmet sich deshalb den folgenden Fragen:

1. Wie gewinnen wir geeignete Formbeschreibungen aus den Ergebnissen des vorangehenden Segmentierungsprozesses?
2. Wie extrahieren oder identifizieren wir Form- und Gestaltmerkmale?
3. Wie bearbeiten wir Form- und Gestaltbeschreibungen?

Die gestaltorientierte Darstellung und Verarbeitung von Bildinformation wird als *morphologische Bildverarbeitung* bezeichnet<sup>1</sup> — unter dem Begriff '*morphologische Bildoperationen*' wird allerdings ausschließlich eine spezielle Klasse von Operationen verstanden, die im Abschnitt 7.3.3 (Seite 191) behandelt wird.

Zunächst sollen einige Grundbegriffe zu einer Geometrie des Bildes vorausgeschickt werden, auf der die Verbindungen benachbarter Pixel zu Konturen und/oder Regionen so dargestellt werden können, dass der Übergang zu geeigneteren Datenstrukturen ermöglicht wird.

---

1. Soille P: Morphologische Bildverarbeitung, Springer Berlin–Heidelberg 1998

## 7.1.1 Die diskrete Geometrie der Pixelwelt

Datenstrukturen, die für eine gestaltorientierte Beschreibung geeignet sind, müssen auf einer geometrischen Beschreibung der Zusammengehörigkeiten von Pixeln basieren. Das Raster der Pixel stellt ein *Diskretisierungsnetzwerk* dar, für dessen geometrische Behandlung eine *diskrete Geometrie* Anwendung findet<sup>1</sup>. Die Pixel werden durch die *Knoten* des Diskretisierungsnetzwerkes repräsentiert.

Der *Graph* definiert Nachbarschaftsbeziehungen zwischen Knoten des Diskretisierungsnetzwerkes. Die Verbindung zwischen zwei Punkten heißt *Kante* (nicht zu verwechseln mit der Kante als Objektberandung!). Spezielle, für diskrete Bilder relevante Graphen sind der einfache Graph und der ebene Graph:

- Ein *einfacher Graph* ist dadurch gekennzeichnet, dass keine Schleifen vorkommen und zwischen zwei Knoten maximal eine Kante existiert.
- Ein *ebener Graph* ist dadurch gekennzeichnet, dass sich Kanten an keiner Stelle kreuzen.

Der Abstand zwischen zwei Knoten ist durch die *Distanzfunktion*, der minimalen Zahl von Kanten zwischen zwei Punkten gegeben. Die Distanzfunktion ist nicht identisch mit dem euklidischen Abstand — siehe auch Abschnitt 8.2, Seite 198!

In dem einfachen Graphen, der ein diskretes zweidimensionales Bild repräsentiert, bestimmt die *Konnektivität* den Freiheitsgrad, unter dem Verbindungen zwischen benachbarten Knoten hergestellt werden können, um Pixel mit gleichem Etikett zu Kontur- oder Regionengraphen zusammen zu schließen. Üblicherweise werden der Behandlung diskreter Bilder die folgenden Konnektivitäten zugrunde gelegt:

---

1. Der Bezug zwischen den Koordinaten der als *Bildpunkte* behandelten kleinsten Teilflächen eines digitalisierten Bildes wurde in Bild 1—16 auf Seite 30 festgelegt.



- **Vierer-Nachbarschaft (4N-Konnektivität)**  
Die Knoten sind in einem orthogonalen Raster angeordnet. Von einem Knoten existieren Kanten nur zu seinem oberen, unteren, linken und rechten Nachbarknoten, Bild 7—1. Der resultierende Graph ist ein einfacher ebener Graph, auch *Schachbrett-* oder *Manhattan-Graph* genannt.
- **Achter-Nachbarschaft (8N-Konnektivität)**  
Die Knoten sind ebenfalls in einem orthogonalen Raster angeordnet. Von einem Knoten existieren außer den Kanten der Vierer-Nachbarschaft auch Kanten zu allen vier diagonalen Nachbarknoten, Bild 7—2. Der resultierende Graph ist ein einfacher nicht-ebener Graph.
- **Sechser-Nachbarschaft (6N-Konnektivität)**  
Sind die Knoten in einem *hexagonalen* Raster angeordnet, wird eine kreuzungsfreie Sechser-Nachbarschaft mit isotroper Kantenlänge ermöglicht, Bild 7—3. Der resultierende Graph ist ein einfacher ebener Graph. Hexagonale Raster werden jedoch trotz einiger Vorteile (siehe unten) kaum eingesetzt, weil zunächst eine geometrische Transformation der üblicherweise gegebenen kartesischen Pixelraster einschließlich der damit verbundenen Interpolationen erforderlich ist.

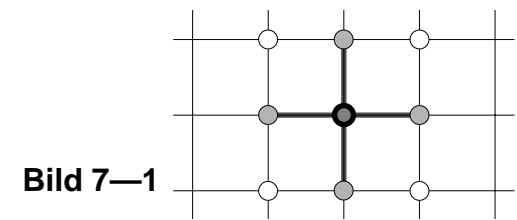


Bild 7—1

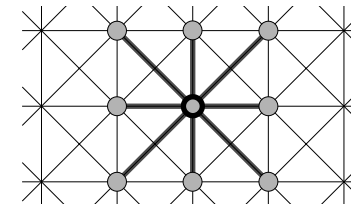


Bild 7—2

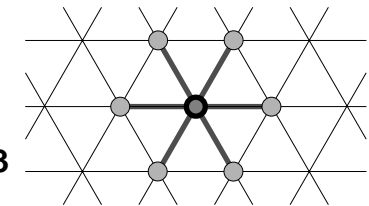


Bild 7—3

Konsistenzprobleme können sich ergeben, wenn sowohl Regionen (einfach zusammenhängende Knotenbereiche) als auch Konturen (Ketten von Knoten, die jeweils durch nur eine Kante verbunden sind) durch Graphen derselben Konnektivität repräsentiert werden sollen: Im Fall einer 8N-Konnektivität können Regionen Konturen an diagonalen Fortsetzungen durchdringen, in Bild 7—4 die durch einen gestrichelten Kreis gekennzeichnete Brücke. Im Fall einer 4N-Konnektivität können Konturen Regionen an diagonalen Einschnürungen zertrennen, in Bild 7—5 die durch einen gestrichelten Kreis gekennzeichnete Barriere. Eine Konsistenz kombinierter Konturen- und Regionen-Graphen ist nur gegeben, wenn Konturen in 8N- und Regionen in 4N-Konnektivität oder umgekehrt repräsentiert werden.

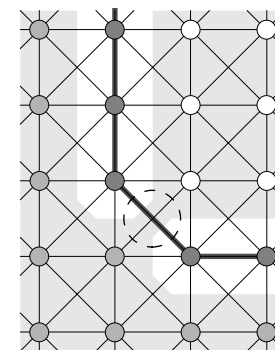


Bild 7—4

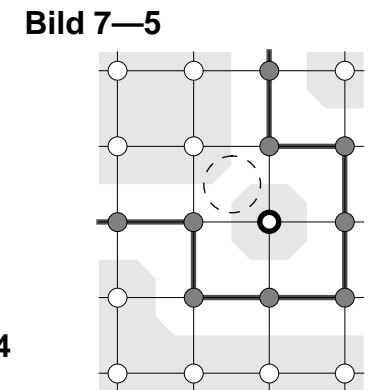


Bild 7—5

Hier besitzt das hexagonale Raster einen wesentlichen Vorteil: Sowohl Konturen als auch Regionen können in 6N-Konnektivität repräsentiert werden, ohne dass sich ein Konsistenzproblem ergibt. Um die algorithmischen Schwierigkeiten hexagonaler Raster zu umgehen, kann die 6N-Konnektivität in einem pseudo-hexagonalen Raster dargestellt werden, Bild 7—6.

Diese Überlegungen sind eine wichtige Grundlage zur Formulierung von Algorithmen, mit denen geschlossene Darstellungen von Konturen oder Regionen hergestellt werden sollen. Unsere 'Vogelperspektive', mit der wir zweidimensionale Bilder aus der dritten Dimension ganzheitlich zu erfassen in der Lage sind, verleitet uns leicht zu einer Unterschätzung der Problematik entsprechender Rechenprozesse: Auf die Pixeldarstellung aufsetzende Rechenprozesse bewegen sich gleichsam wie Käfer in einer diskreten, zweidimensionalen Landschaft und können dabei lediglich den Zustand ihrer nächsten Nachbarorte ertasten, um Entscheidungen für ihre nächste Schritte zu fällen. Dazu brauchen sie konsistente Regeln.

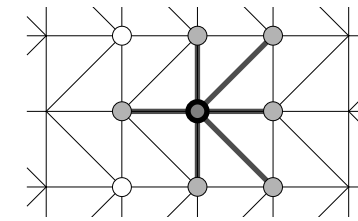


Bild 7—6

## 7.2 Kontinuierliche und diskrete Konturbeschreibungen

Die Ausgangssituation für die Gewinnung einer geschlossenen Darstellung einer Kontur wird im allgemeinen ein Pixelfeld mit etikettierten Pixeln sein. Es gilt nun, diese sehr schlecht für eine Analyse der Kontur-Gestaltmerkmale und deren Bearbeitung geeignete Darstellung in besser geeignete Beschreibungsklassen umzusetzen.

Zwar werden für alle anschließenden Rechenprozesse wiederum diskrete Beschreibungen benötigt. Die Eigenschaften von Konturen und deren Beschreibungen lassen sich jedoch übersichtlicher im Kontinuierlichen betrachten. Bevor auf die spezielle Problematik der diskreten Beschreibungen eingegangen wird, sollen daher zunächst kontinuierliche Konturbeschreibungen, aus ihnen ableitbare Gestaltmerkmale und einige spezielle Analyseverfahren behandelt werden.

7.2.1 Kontinuierliche Konturbeschreibungen

Die Gewinnung von Gestaltmerkmalen dient unterschiedlichen Zielen:

- Objekte sollen auf der Basis ihrer Ähnlichkeit zu einfachen geometrischen Figuren (z.B. Kreis, Ellipse) klassifiziert werden.
- Objekte sollen auf der Basis einfacher geometrischer Merkmale (z.B. Hauptachsenlage, Konturlänge, charakteristische Krümmungen) klassifiziert werden.
- Konturen oder Konturabschnitte sollen auf Grund ihrer Verläufe verändert, ergänzt, zusammengefügt, getrennt oder verworfen werden.

Es kann nun versucht werden, gegebene Konturen ganz oder in Abschnitten auf geometrische Funktionen zurück zu führen. Am Beispiel eines Kreises um den Ursprung (Bild 7—7) lassen sich drei Alternativen für die formale Darstellung einer geometrischen Figur angeben:

- die explizite Darstellung, Gl. (7-1),
- die implizite Darstellung, Gl. (7-2) und
- die parametrische Darstellung, Gl. (7-3).

Für eine Reihe von Ansätzen zur Gewinnung von Gestaltmerkmalen und Verfahren zur Konturbearbeitung sind parametrische Darstellungen am besten geeignet. So lassen sich von der Darstellung einer gegebenen Kontur über dem Konturweg  $s$  leicht die den Konturverlauf beschreibenden Merkmale *Richtung*  $\varphi(s)$  und *Krümmung*  $K(s)$  ableiten. Benötigt werden dafür die partiellen Ableitungen  $x'$  und  $y'$  bzw.  $x''$  und  $y''$ , Gl. (7-4).

Die Konturrichtung ergibt sich nach Gl. (7-5). Die Richtungsfunktion geschlossener Konturen wird üblicherweise als so genanntes  $\varphi$ - $s$ -Diagramm oder seine Variante, das  $\psi$ - $s$ -Diagramm dargestellt, Gl. (7-6). Das  $\psi$ - $s$ -Diagramm beschreibt die Abweichung der Kontur von einem Kreis. Die Funktion ist über der Gesamtlänge  $S$  der Kontur periodisch. Beide Darstellungen sind invariant gegenüber einer Translation des Objektes. Eine Skalierung

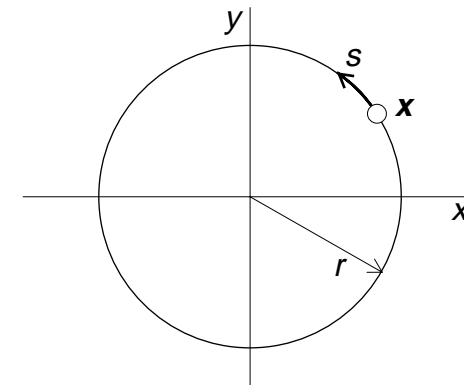


Bild 7—7

$$y = f(x) \qquad y = \sqrt{r^2 - x^2} \qquad (7-1)$$

$$0 = F(x,y) \qquad 0 = x^2 + y^2 - r^2 \qquad (7-2)$$

$$\vec{x}(s) = \begin{Bmatrix} x(s) \\ y(s) \end{Bmatrix} \qquad \begin{matrix} x = r \cos 2\pi s \\ y = r \sin 2\pi s \end{matrix} \qquad (7-3)$$

$$x' = \frac{\partial s}{\partial x}; \quad y' = \frac{\partial s}{\partial y} \qquad (7-4)$$

$$\varphi = f(s) = \operatorname{atan} \frac{y'}{x'} \qquad (7-5)$$

$$\psi(s) = \varphi(s) - 2\pi(s/S) \qquad (7-6)$$

des Objektes führt zu einer entsprechenden Skalierung der  $s$ -Achse, eine Rotation zu einer Verschiebung auf der  $s$ -Achse. Invarianz gegenüber Skalierung und Rotation kann durch eine Normierung erzielt werden. Bild 7—8 zeigt das  $\varphi$ - $s$ - und das  $\psi$ - $s$ -Diagramm einer einfachen Beispielkontur.

Abschnitte starker Steigung des  $\varphi$ - $s$ -Diagrammes markieren Ecken einer Kontur, die Projektion dieser Abschnitte auf die  $\varphi$ -Achse geben deren Winkel an. Abschnitte geringer Steigung kennzeichnen annähernd gerade Konturabschnitte, deren Länge aus der Projektion auf die  $s$ -Achse abzulesen ist.

Interessante Rückschlüsse können aus dem Histogramm von  $\varphi$  über  $s$ , der so genannten *Shape-density-Funktion* gewonnen werden (Bild 7—9 für die Beispielkontur). Überwiegend gerade Konturabschnitte bilden sich als Maxima ab. Ihre Lage auf der  $\varphi$ -Achse ist ein Maß für die Winkel, die sie miteinander bilden. Bereiche auf der  $\varphi$ -Achse mit konstanter Häufigkeit ergeben sich bei Konturabschnitten mit konstantem Kurvenradius.

Ein weiteres interessantes Merkmal ist der Verlauf der Krümmung über dem Konturverlauf, Gl. (7-7). In Bild 7—10, der Segmentierung einer Muskelzellen-Biopsie, dient die Krümmungsinformation der Trennung verschmolzener Objekte.

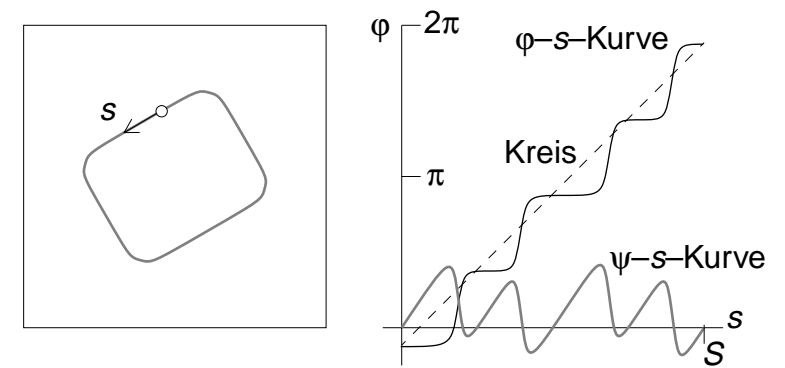


Bild 7—8

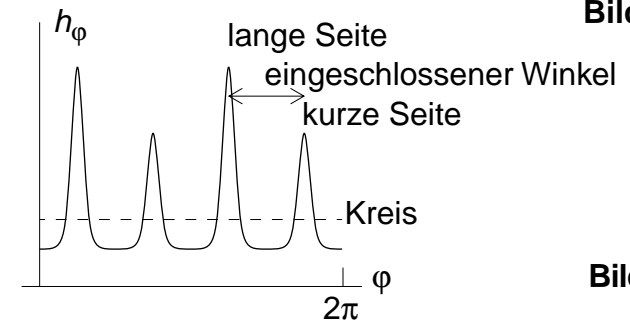


Bild 7—9

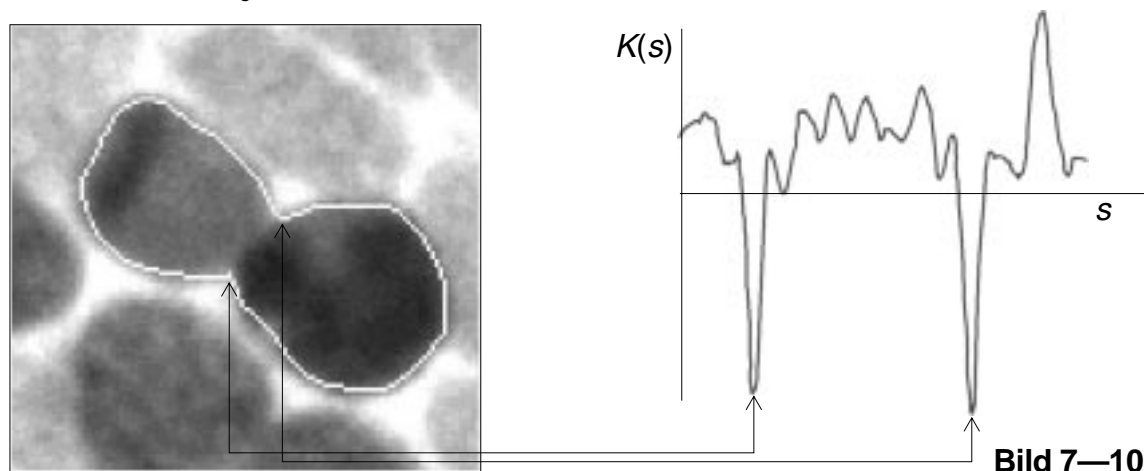


Bild 7—10

$$K = \frac{d\varphi}{ds}$$

$$ds = \sqrt{dx^2 + dy^2}$$

$$K = \frac{x'y'' - x''y'}{\sqrt{x'^2 + y'^2}^3} \quad (7-7)$$

## 7.2.2 Fourier-Deskriptoren

Die Periodizität der Darstellungen von Konturmerkmalen über dem Konturverlauf weckt die Assoziation mit periodischen Zeitsignalen und legt damit die Anwendung der im Zeitbereich geläufigen Signalanalysemethoden nahe. Interessante Aufschlüsse über den Signalcharakter liefert insbesondere die *Fourier-Analyse*. Ohne weiteres können beispielsweise  $\psi$ - $s$ -Funktionen einer Fourier-Transformation unterzogen werden.

Ein eleganterer Weg eröffnet sich, wenn die  $x,y$ -Ebene als komplexe Ebene interpretiert wird. Die Kontur stellt dann eine komplexe Funktion der Bogenlänge  $s$  dar. Ihre Koordinaten  $x,y$  werden Realteil und Imaginärteil der komplexen Konturfunktion, Gl. (7-8). Die geschlossene Kontur wird eine bezüglich des Umfangs  $S$  periodische Funktion. Vorausgesetzt wird, dass für die Kontur die Annahme eines 'bandbegrenzten' Musters gilt, d.h. Konturkrümmungen sind nur insoweit zulässig, als sie sich durch eine endliche Zahl  $I$  von Spektralkoeffizienten  $c_i$  darstellen lassen, Gl. (7-9).

Wie in der Signalanalyse charakterisieren die Spektralkoeffizienten  $c_i$  hier den Konturverlauf. Sie werden deshalb als *Fourier-Deskriptoren* bezeichnet: Koeffizienten niedriger Wellenzahlen sind ein Maß für die Abweichung von einem kreisförmigen Verlauf, Koeffizienten höherer Wellenzahlen kennzeichnen nicht-glatte Konturverläufe.

Interessant an diesem Ansatz ist die Darstellung geometrischer Variationen von  $\underline{x}(s)$  gegenüber einem Referenzmuster  $\underline{x}_R(s)$ . Gl. (7-10) fasst die zulässigen Variationen zusammen:

Translation um  $\underline{x}_0 = x_0 + jy_0$ ,

Rotation mit dem Winkel  $\Phi$  um den Bezugspunkt  $\underline{x}_M$ ,

Aufpunktverschiebung um  $s_0$ ,

Radiale Streckung bezüglich  $\underline{x}_M$  um den Faktor  $a$  (Skalierung)<sup>1</sup>.

$$\underline{x} = \text{Re}(\underline{x}(s)) + j\text{Im}(\underline{x}(s)) = x(s) + jy(s) \quad (7-8)$$

$$\underline{x}(s) = \sum_{i=0}^I c_k e^{j(i\omega s)}; \quad \omega = \frac{2\pi}{S} \quad (7-9)$$

$$\underline{x}(s) = a \cdot \underline{x}_R(s + s_0) \cdot e^{j\Phi} + \underline{x}_0; \quad \omega = \frac{2\pi}{aS} \quad (7-10)$$

1. Die Bogenlänge ist über die Skalierung der 'Kreisfrequenz'  $\omega$  normiert.

Damit ergeben sich für die Koeffizienten nach Gl. (7–11) die folgenden Bedeutungen:<sup>1</sup>

- $c_0$  gibt die Lage des Schwerpunktes an; sie ist invariant gegenüber Aufpunktverschiebungen.
- Das Amplitudenspektrum, d.h. die Menge aller Koeffizientenbeträge  $|c_{i \neq 0}|$  stellt einen vollständigen Satz Lage-invarianter Merkmale des Konturverlaufs dar.
- $|c_{i \neq 0}|$  ist eine Funktion der Streckung  $a$ ; sie ist invariant gegenüber einer Translation  $\underline{x}_0$ .
- Das Argument von  $c_{i \neq 0}$  ist eine Funktion des Aufpunktes  $s_0$  und des Rotationwinkels  $\Phi$ .

Mit entsprechenden Normierungen kann Invarianz gegenüber Rotation und/oder Skalierung (Streckung) hergestellt werden.

### 7.2.3 Diskrete Konturbeschreibung

Ist eine Kontur als Kette benachbarter etikettierter Pixel gegeben, kann leicht eine Koordinaten-Liste aller Kontur-Pixel aufgestellt werden. Die primitivste Form einer parametrischen Darstellung entsteht, wenn die Liste entsprechend der örtlichen Verkettung der Kontur-Pixel geordnet ist. Der Parameter ist der diskrete Weg in Inkrementen der Distanz 1. Bei einer 8N-Konnektivität kann die euklidische Distanz somit 1 oder  $\sqrt{2}$  sein.

Weniger redundant und für die Analyse und Bearbeitung besser zugänglich ist der *Kettenkode* (*chain code*, *Freeman code*): Von den Koordinaten des Aufpunktes ausgehend wird mittels einer Liste von Richtungsdeskriptoren lediglich die Fortsetzungsrichtung der Kontur beschrieben. Je nach Konnek-

$$\begin{aligned} c_0 &= c_{R0} \cdot a \cdot e^{j\Phi} + \underline{x}_0 \\ c_i &= c_{Ri} \cdot a \cdot e^{j(\Phi + is_0)} + \underline{x}_0; \quad i \neq 0 \end{aligned} \quad (7-11)$$

1. Burkhardt H: Transformationen zur lageinvarianten Merkmalsgewinnung (Habilitationsschrift), VDI-Fortschritt-Berichte Reihe 10, Nr. 7, VDI-Verlag Düsseldorf 1979

tivität ist die Fortsetzungsrichtung in vier oder acht Stufen zu kodieren. Der Kettenkode ermöglicht einfache Operationen wie die Bestimmung der Konturlänge, der Fläche einer umschlossenen Region oder einfacher Gestaltcharakteristika (Konvexität, Konkavität etc.).

Für die Bestimmung von Formparametern wie Konturrichtung oder –krümmung ist der Kettenkode auf Grund seiner sehr groben Winkelauflösung ungeeignet. Wird angenommen, dass eine Pixel-Kette die diskrete Darstellung einer eher glatten Kontur sei, können solche Parameter präziser aus einer in größeren Intervallen abgetasteten Darstellung ermittelt werden.

Die Abtastung an  $I$  Konturpunkten  $s_j$  lässt sich als Vektorpaar aus Konturpixel-Koordinaten schreiben, Gl. (7-12).<sup>1</sup> Geläufig ist für diese Darstellung der Begriff *finite Differenzen*. Wichtig sind die Ableitungen:

- Die 1. Ableitung, der symmetrische Gradient nach Gl. (7-13), entspricht bei äquidistanter Abtastung dem Richtungsvektor im Konturpunkt  $s_j$ . Bei nicht-äquidistanter Abtastung ist eine Normierung erforderlich.
- Die 2. Ableitung, Gl. (7-14), ist Ausgangspunkt für die Berechnung der Krümmung. Sie kann als diskrete Faltung der Konturvektoren formuliert werden, Gl. (7-15).

Die Krümmung kann als  $z$ -Komponente des Kreuzproduktes aus 1. und 2. Ableitung dargestellt werden, Gl. (7-16). Abgesehen von einer Normierung kann in Gl. (7-16) die diskrete Formulierung von Gl. (7-7), Seite 180 erkannt werden. Die Notwendigkeit einer Normierung entfällt bei einer äquidistanten Abtastung der Kontur. Eine Näherung für geringe Krümmungen ergibt das Kreuzprodukt aus den Vorwärtsgradienten  $\Delta^+ x$ , Gl. (7-17).

1. Der Begriff Vektor wird wie üblich nicht nur für Ortsvektoren sondern auch für die Bezeichnung eindimensionaler Datenfelder verwandt. Zu den Konventionen für die Verwendung von Klammern siehe Abschnitt 1 (Fußnote auf Seite 16).

$$\langle x(s) \rangle = \langle x(s_0) x(s_1) \dots x(s_I) \rangle \tag{7-12}$$

$$\langle y(s) \rangle = \langle y(s_0) y(s_1) \dots y(s_I) \rangle$$

$$\overrightarrow{\Delta x}(s_j) = \begin{Bmatrix} \Delta x(s_j) \\ \Delta y(s_j) \end{Bmatrix} = \frac{1}{2} \begin{Bmatrix} x(s_{i+1}) - x(s_{i-1}) \\ y(s_{i+1}) - y(s_{i-1}) \end{Bmatrix} \tag{7-13}$$

$$\overrightarrow{\Delta_2 x}(s_j) = \begin{Bmatrix} \Delta_2 x(s_j) \\ \Delta_2 y(s_j) \end{Bmatrix} = \frac{1}{4} \begin{Bmatrix} x(s_{i+1}) - 2x(s_i) + x(s_{i-1}) \\ y(s_{i+1}) - 2y(s_i) + y(s_{i-1}) \end{Bmatrix} \tag{7-14}$$

$$\Delta_2 x(s_j) = x(s_j) * \frac{1}{4} [1 \ -2 \ 1] \tag{7-15}$$

$$\vec{K}(s_j) = \overrightarrow{\Delta_2 x}(s_j) \times \overrightarrow{\Delta x}(s_j) \tag{7-16}$$

$$\tag{7-17}$$

$$K(s_j) = \Delta^+ x(s_j) \cdot \Delta^+ y(s_{j-1}) - \Delta^+ x(s_{j-1}) \cdot \Delta^+ y(s_j)$$

$$\begin{Bmatrix} \Delta^+ x(s_j) \\ \Delta^+ y(s_j) \end{Bmatrix} = \begin{Bmatrix} x(s_{i+1}) - x(s_i) \\ y(s_{i+1}) - y(s_i) \end{Bmatrix}$$

### 7.2.4 Approximation einer kontinuierlichen Kontur

Die Diskretisierung einer Kontur mittels finiter Differenzen hat den Vorteil einer einfachen Implementierung. Nachteilig ist jedoch, dass die Approximation von Ableitungen, wie sie für die Darstellung der Konturnormalen oder der Konturkrümmung benötigt werden, ungenau ist. Nachteilig ist auch, dass eine Überabtastung erforderlich ist, wenn die Kontur im Verlauf ihrer Bearbeitung gedehnt werden soll. Diesen Nachteilen wird durch Konturdarstellungen mit abschnittsweise kontinuierlicher Approximation begegnet. Zwei unterschiedliche Ansätze hierfür sind geläufig: *Splines* und *Finite-Elemente-Modelle (FEM)*.

Splines sind Funktionen, die sich wie eine elastische Latte an eine Folge von Stützstellen anschmiegen. In Abschnitt 5.6 auf Seite 131 wurde bereits ihre Anwendung für die Interpolation der Grauwertfunktion beschrieben. Kubische B-Splines beispielsweise sind Polynome dritten Grades, die sich nach beiden Seiten des zu approximierenden Kurvenpunktes auf jeweils zwei benachbarte Abtastwerte stützen.

Ebenfalls durch Polynome approximieren Finite-Elemente-Modelle eine Kontur, im Gegensatz zu Splines jedoch jeweils innerhalb *finiter* Konturabschnitte, d.h. die Approximation stützt sich allein auf die Daten der beiden jeweils unmittelbar benachbarten Abtastwerte. Die Approximationsgüte ist dennoch vergleichbar mit einer Spline-Approximation, da auch bei den FEM die Fortsetzung der Funktion über das betrachtete Intervall hinaus berücksichtigt wird — hier jedoch nicht in Form weiterer Stützstellen, sondern auf der Basis der Ableitungen an den Intervallgrenzen. FEM bieten Vorteile für die numerische Behandlung, insbesondere bei nicht-äquidistanter Abtastung der Kontur. Sie sollen deshalb im folgenden detaillierter beschrieben werden.

Der Abschnitt  $S$  einer Kontur  $\mathbf{x}(s)$ , auf dem der Laufindex  $s$  definiert ist, wird in  $I$  Intervalle  $[s_{i-1}, s_i]; [s_i, s_{i+1}]$  aufgeteilt. Über jedem dieser Intervalle



wird der Kurvenverlauf jeweils durch ein Polynom dargestellt, z.B. durch ein kubisches Polynom nach Gl. (7-18)<sup>1</sup>. Für die Darstellung des gesamten Konturabschnitts  $S$  sind also  $I$  Vektoren von jeweils vier polynominalen Entwicklungskoeffizienten  $\langle a_0 \ a_1 \ a_2 \ a_3 \rangle^T$  nötig.

Eine wichtige Eigenschaft der FEM ist die Stetigkeit an den Intervallgrenzen. Diese erhält man, wenn man den Konturverlauf nach Polynomen entwickelt, bei denen jeweils nur eine der Größen  $x(s_j)$ ,  $x'(s_j)$ ,  $x(s_{i+1})$ ,  $x'(s_{i+1})$  von Null verschieden ist. Für ein Intervall  $s = [0,1]$  sind dies die vier Polynome  $p_{1...4}$  nach Gl. (7-19), dargestellt in Bild 7-11. Für andere Intervallgrenzen  $[s_{i-1}, s_i]$  enthält man die entsprechenden Basisfunktionen durch Dehnung bzw. Stauchung und Verschiebung.

Die Approximation nach Gl. (7-18) hat die Eigenschaft  $\hat{x}(s_j)=q_1$ ,  $\hat{x}'(s_j)=q_2$ ,  $\hat{x}(s_{i+1})=q_3$ ,  $\hat{x}'(s_{i+1})=q_4$ . Das heißt, die Werte der Approximation und ihre Ableitungen an den Intervallgrenzen legen den Funktionsverlauf innerhalb des Intervalls eindeutig fest, Gl. (7-20). Die vier Entwicklungskoeffizienten der kubischen Polynome entsprechen dabei den vier Freiheitsgraden der Funktionswerte und der ersten Ableitungen an den Intervallgrenzen. Wenn  $x(s_j)$  und  $x'(s_j)$  den Funktionsverlauf in den beiden angrenzenden Intervallen  $[s_j, s_{i+1}]$  und  $[s_{i-1}, s_j]$  bestimmen, ergibt sich die gewünschte  $C^1$ -Stetigkeit, d.h. sowohl der Funktionswert als auch die Steigung der Konturintervalle schließen aneinander an.

Der gesamte Funktionsverlauf der Approximation über den Abschnitt  $S$  ist somit durch die Wertepaare  $\hat{x}(s_j)$  und  $\hat{x}'(s_j)$  an den beiden Intervallgrenzen, den *Knoten*  $s_i$  ( $i = 1 \dots I$ ) beschrieben. Die Wertepaare  $\hat{x}(s_j)$ ,  $\hat{x}'(s_j)$  werden daher Knotenvariable  $q_{1...4}$  genannt. Soll nun die Koordinate eines beliebigen Kurvenpunktes  $s_{i-1} \leq s < s_i$  ermittelt werden, bestimmen die Funktions-

$$\hat{x}(s) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 = \langle 1 \ x \ x^2 \ x^3 \rangle \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{Bmatrix} \quad 0 \leq s < 1 \quad (7-18)$$

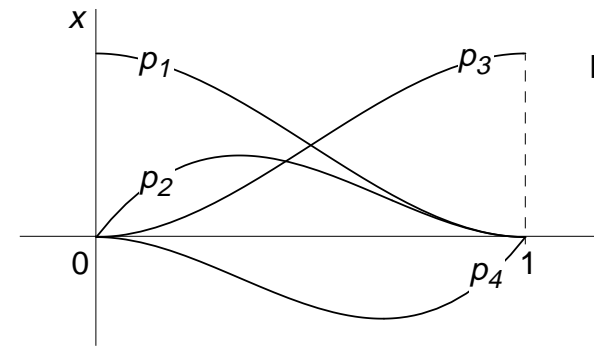


Bild 7-11

$$(7-19) \quad \begin{aligned} p_1(s) &= 1 - 3s^2 + 2s^3 \\ p_2(s) &= s - 2s^2 + s^3 \\ p_3(s) &= 3s^2 - 2s^3 \\ p_4(s) &= -s^2 + s^3 \end{aligned}$$

$$\hat{x}(s) = \langle p \rangle \{ q \} = \langle p_1(s) \ p_2(s) \ p_3(s) \ p_4(s) \rangle \begin{Bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{Bmatrix} \quad s_{i-1} \leq s < s_i \quad s = \frac{s - s_{i-1}}{s_i - s_{i-1}} \quad (7-20)$$

1. Der besseren Übersichtlichkeit halber werden die folgenden Gleichungen nur für die  $x$ -Koordinate aufgestellt. Die entsprechende Formulierung gilt jeweils auch für die  $y$ -Koordinate.

werte der vier Funktionen  $p_{1...4}$  an der Position  $s-s_{i-1}$ , mit welchem Gewicht die vier Knotenvariablen der Intervallgrenzen in das Ergebnis eingehen. Die Ermittlung der Koordinate kann somit als Faltung der Abtastfunktion mit zwei sogenannten *Formfunktionen* aufgefasst werden, die aus den Polynompaaren  $p_1, p_2$  und  $p_3, p_4$  so konstruiert sind, dass sie sich jeweils über eine Intervall-Länge  $2(s_i-s_{i-1})$  erstrecken, Bild 7—12.

Günstig für die weitere numerische Behandlung kann auch eine Matrix-Formulierung sein. Dazu werden jeweils Formfunktionenpaare  $p_i, p'_i$ , die sich über den gesamten Konturabschnitt  $S$  erstrecken, für alle Intervalle definiert, Gl. (7-21). Werden diese zu einem Vektor zusammengefasst, kann die Approximation über den Konturabschnitt  $S$  als Matrixgleichung (7-22) geschrieben werden.  $\langle P \rangle$  ist darin der Zeilenvektor der Formfunktionen und  $\{Q\}$  der Spaltenvektor der Knotenvariablen.

Durch die  $2I$  Stetigkeitsbedingungen an den Intervallgrenzen wird die Zahl der Konturfreiheitsgrade von  $4I$  auf  $2I$  reduziert. Die Ableitungen der Kontur können nun sehr elegant auf den Formfunktionen ausgeführt werden, Gl. (7-23).

Mit den verschiedenen parametrischen Darstellungen können nunmehr Gestalteigenschaften einer Kontur<sup>1</sup> analysiert und formuliert werden. Für eine nachfolgende Klassifikation können Gestaltmerkmale hergeleitet werden wie etwa Glattheit, Rauigkeit oder Welligkeit einer Kontur oder ihre Ähnlichkeit zu Referenzfiguren. Darüber hinaus liefern insbesondere die approximativen Verfahren Ansätze für eine Kontursuche oder -anpassung nach Gestaltkriterien, wie beispielsweise die unter dem Begriff *aktive Konturen* (siehe Seite 190) zusammengefassten Methoden.

1. Wie auf Konturen können die beschriebenen Verfahren selbstverständlich auch auf linienhafte Strukturen angewandt werden, die beispielsweise mittels einer Skelettierung (siehe Seite 193) gewonnen worden sind.

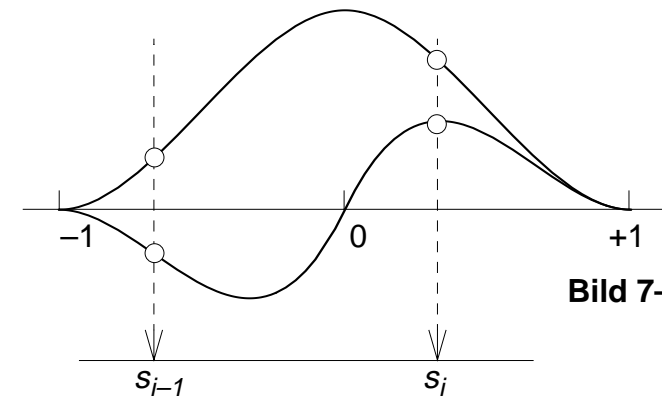


Bild 7—12

$$p_i(s) = 1 - 3s^2 + 2|s|^3$$

$$p'_i(s) = s - 2s \cdot |s| + s^3$$

$$p_i, p'_i = 0 \quad \text{außerhalb } [s_i, s_{i+1}] \quad (7-21)$$

$$\hat{x}(s)|_0^S = \langle P \rangle \{Q\} \quad (7-22)$$

$$\{Q\} = \langle \hat{x}(s_1) \hat{x}'(s_1) \dots \hat{x}(s_i) \hat{x}'(s_i) \dots \hat{x}(s_I) \hat{x}'(s_I) \rangle^T$$

$$\langle P \rangle = \langle p_1(s) p'_1(s) \dots p_i(s) p'_i(s) \dots p_I(s) p'_I(s) \rangle$$

$$\begin{aligned} \frac{\partial}{\partial s} \hat{x}(s) &= \left\langle \frac{\partial P}{\partial s} \right\rangle \{Q\} \\ \frac{\partial^2}{\partial s^2} \hat{x}(s) &= \left\langle \frac{\partial^2 P}{\partial s^2} \right\rangle \{Q\} \end{aligned} \quad (7-23)$$

### 7.3 Mit Gestaltkriterien auf Kontursuche

Warum gibt man dem Pilzesammel–Novizen den Tip, hin und wieder auf bereits gefundene Pilze im Sammelkorb zu schauen? Warum entdecken wir bekannte Objekte schlecht, wenn sie in einer unerwarteten Orientierung oder Umgebung erscheinen — ein aus dem ‘Suchbild’ bekannter Effekt?

Für den Wahrnehmungsprozess spielt offensichtlich der Vergleich zwischen aktuellen visuellen Eindrücken und mentalen Bildern eine zentrale Rolle („wir sehen nichts, was wir nicht bereits kennen“). Dieser Vergleich findet sicher nicht auf der Basis von Pixelfeldern statt. Verglichen werden vermutlich bereits extrahierte Strukturen und Merkmale auf Abstraktionsebenen, auf denen leicht Invarianz gegenüber affinen Abbildungen, Beleuchtungssituationen etc. hergestellt werden kann.

Die Ebene der Konturbilder hat hier besondere Bedeutung, wie die unmittelbare Assoziation einer Grafik mit der dargestellten Realität oder die Wiedererkennung einer bekannten Person aus einer Karikatur belegen. Für die Bildanalyse stellen sich die Fragen:

1. Wie finden wir in gestörten Bildern geometrische Grundstrukturen, aus denen sich das gesuchte Objekt aufbaut?
2. Wie passen sich Konturen oder Linienstrukturen mit vorgegebenen Gestaltmerkmalen optimal in eine gegebene Pixel–‘Landschaft’ ein?
3. Wie können Bildinhalte auf der Gestaltebene so manipuliert werden, dass sie mit den zugrunde gelegten Gestaltmodellen korrespondieren?

Für diese Fragestellungen sollen im folgenden beispielhafte Konzepte beschrieben werden.

## 7.3.1 Die Hough-Transformation

Visuelle Experimente wie das Beispiel in Bild 6—13 auf Seite 155 zeigen, dass Objekte von einfacher Geometrie wie glatte Linien, Kreise oder Quadrate selbst in sehr stark gestörten Szenen entdeckt werden. Dies legt die Vermutung nahe, dass über lokale Strategien hinaus wie z.B. der Suche nach der 'guten Fortsetzung' (Seite 174) in globalerer Weise nach Linienstrukturen, die einfachen geometrischen Figuren folgen, gesucht wird, indem das Bildsignal entlang solcher Trajektorien gleichsam aufintegriert wird.

Die maschinelle Umsetzung dieser Hypothese führt zu einer Suche nach Strukturelementen (englisch 'primitives') wie Geradenabschnitten oder Kreisen nach der Maßgabe, "wo passen sich die Pixel z.B. eines Gradientenbildes an das vorgegebene Element an?" Die *Hough-Transformation*<sup>1</sup> setzt diesen Gedanken um. Ihr Ausgangspunkt ist, dass die in Betracht gezogenen Strukturelemente durch nur wenige Parameter in ihrer expliziten bzw. impliziten Darstellung bestimmt werden. Die Transformation besteht darin, für die bestimmenden Parameter einen Parameterraum zu definieren. In diesem Raum wird das Strukturelement durch einen *Punkt* repräsentiert. Die Dimension des Parameterraumes wird durch die Zahl der das Strukturelement bestimmenden Parameter festgelegt.

Die Hough-Transformation soll an dem einfachsten Strukturelement, der Geraden erläutert werden. Die Geradengleichung (7-24) im  $x,y$ -Bildraum führt zu einem Parameterraum mit den Koordinaten  $a,b$ , Bild 7—13 oben. Für die Hough-Transformation gilt die Reziprozität: Ein Punkt im Bildraum transformiert sich in ein Strukturelement derselben Klasse im Parameterraum, Bild 7—13 mitte. Für die Gerade wird die inverse Hough-Transformation durch Gl. (7-25) realisiert.

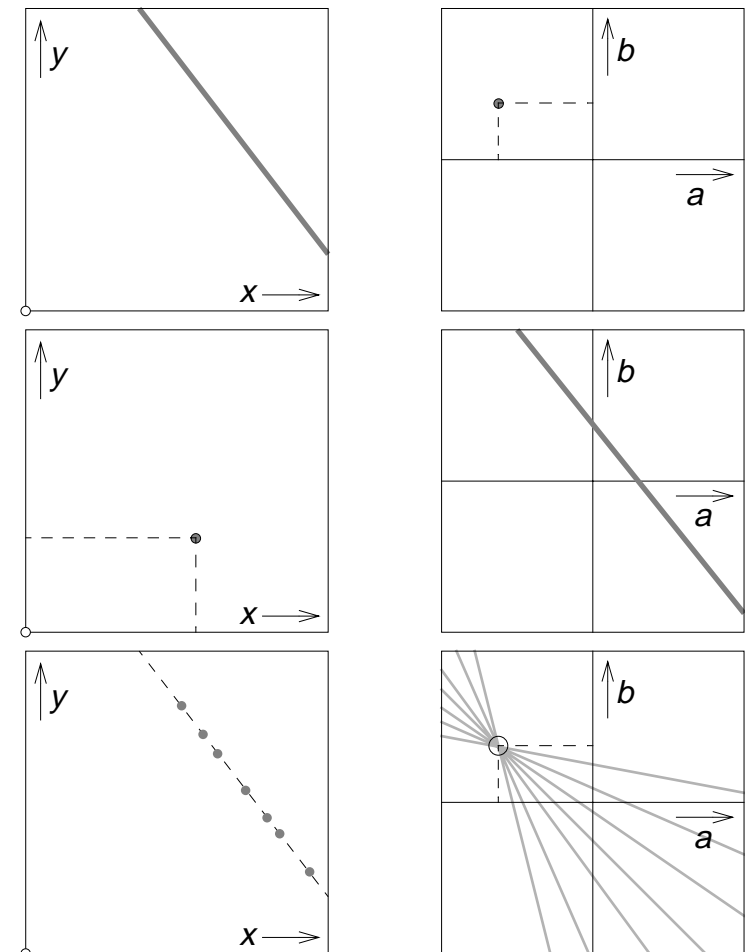


Bild 7—13

1. Leavers VF: Shape Detection in Computer Vision Using the Hough Transform, Springer London 1992  
Schalkoff RJ: Digital Image processing and Computer Vision, John Wiley & Sons New York 1989, S 295—303

$$y = ax + b \quad (7-24)$$

$$b = -xa + y \quad (7-25)$$

Die Hough-Transformierte eines Bildes wird nun erzeugt, indem für jedes Pixel  $g(x,y)$  das gewählte Strukturelement — hier also die Gerade  $b = -xa + y$  — in der zur Pixelposition  $x,y$  korrespondierenden Lage mit der jeweiligen Pixel-Amplitude  $g$  in den Parameterraum eingetragen wird. Wenn sich im Bildraum eine Gerade  $y = ax + b$  abzeichnet, führt die Überlagerung aller Geraden im Parameterraum zu besonders hohen Amplituden an der korrespondierenden Position  $a,b$ , Bild 7—13 unten. Damit wird eine Suche nach dem ausgewählten Strukturelement im Bildraum zur Suche nach lokalen Verdichtungen im Parameterraum.

Die oben gewählte Geraden-Darstellung wird allerdings bei Geraden mit nahe vertikaler Ausrichtung problematisch. Geeigneter ist eine Formulierung mit den Parametern Winkellage  $\phi$  und Ursprungsabstand  $r$ ; Gl. (7-26). Die Position  $x,y$  im Bildraum legt in diesem Fall Amplitude und 'Phasenlage' einer Cosinus-Schwingung im Parameterraum fest, Gl. (7-27). Geradenabschnitte ergeben fächerartige Verdichtungen im Parameterraum, da der Koordinatenbereich der zu einer Position im Parameterraum korrespondierend Bildraum-Geraden nicht vollständig ausgeschöpft wird, Bild 7—14. Bild 7—15 zeigt die Transformierte des Gradientenbildes eines Objektes.

Die Hough-Transformation hat sich als gut geeignet für Parameterräume bis etwa zur vierten Dimension erwiesen. Als Strukturelemente kommen außer Geradenabschnitten Kreise, Ellipsen, Parabel- oder Hyperbeläste infrage. Bei einer höheren Dimensionalität des Parameterraumes ist es praktisch, wenn der Suchraum durch Vorgabe erwarteter Parameterbereiche eingeschränkt werden kann, z.B. bei Kreisen ungefähr bekannter Größe beliebige Mittelpunktkoordinaten aber eingeschränkter Radiusbereich.

Die Akkumulation im Parameterraum wird zweckmäßigerweise rückwärts gerichtet durchgeführt ('destination driven', siehe "Monadische Bildoperationen" auf Seite 103), d.h. für jede Position im diskreten Parameterraum wird im Bildraum eine Summation der Funktionswerte  $g$  aller Pixel des Strukturelementes — z.B. der korrespondierenden Geraden — durchgeführt.

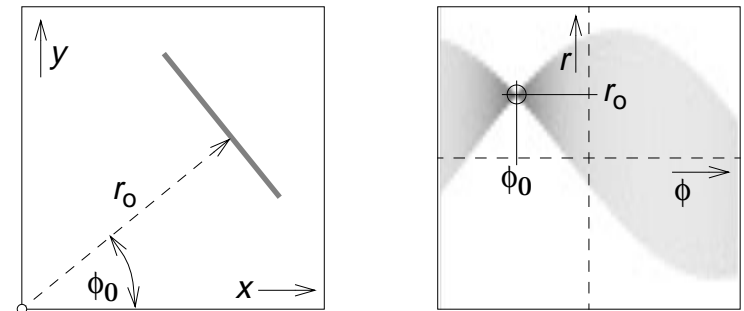


Bild 7—14

$$x \cos \phi + y \sin \phi = r \quad (7-26)$$

$$r = A \cos(\phi + \phi_0) \quad (7-27)$$

$$\phi_0 = \text{atan} \frac{x}{y}; \quad A = \sqrt{x^2 + y^2}$$

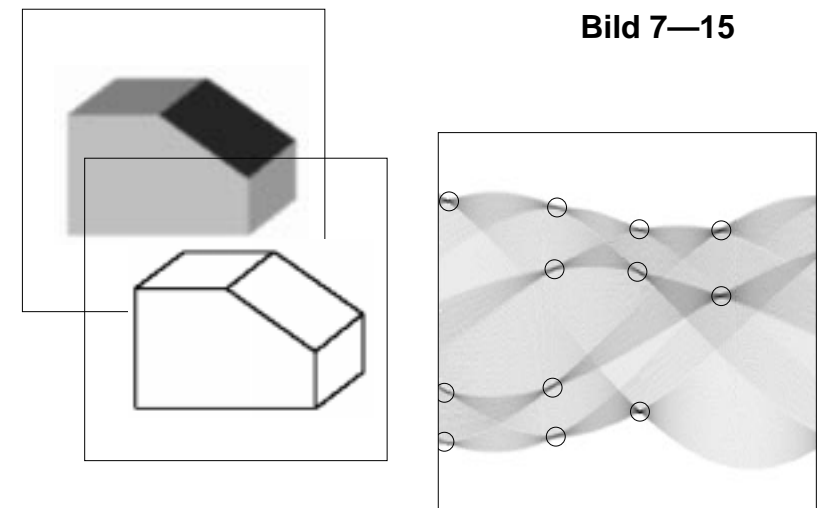


Bild 7—15

## 7.3.2 Aktive Konturen

Die vielfältigen Bezüge zwischen visueller Wahrnehmung und maschineller Bildanalyse legen immer wieder Assoziationen zu Modellen aus unserer Erfahrungswelt nahe, so zum Beispiel das Bild eines Geländes, in dem die Topologie die Bildsignalfunktion abbildet: Kanten werden zu Geländestufen (Bild 6—19 auf Seite 158), lokale Grauwertextrema zu Hügeln oder Tälern, linienförmige Strukturen zu Dämmen oder Gräben.

Wenn jetzt die Aufgabe zu lösen ist, die Lage einer Kontur so festzulegen, dass der beste Kompromiss zwischen einer Anpassung an die Bilddaten und einer angenommenen Glattheit hergestellt wird, kann das Graben-Analogon auf ein pseudomechanisches Modell der Kontur führen: Konturen werden als Masse-behaftete deformierbare Linien behandelt, die sich aufgrund ihrer Schwere in die Gräben legen möchten. Sie können dabei den Windungen und Krümmungen der Gräben nur so weit folgen, wie es ihre Biegesteifigkeit und Längselastizität zulässt. In der englisch sprachigen Literatur wurde für dieses Modell der Begriff 'active contours'<sup>1</sup> eingeführt oder noch anschaulicher das Bild der 'snakes' geprägt.

Das pseudomechanische Modell besteht nun in der Annahme innerer und äußerer Energien:

Mit einer inneren Energiedichte  $e_I$  setzt sich die Kontur einer Verformung entgegen. Mit ihrer Elastizität in Längsrichtung  $\alpha$  und ihrer Biegesteifigkeit  $\beta$  ergibt sich die Modellbeziehung Gl. (7-28), Bild 7—16.

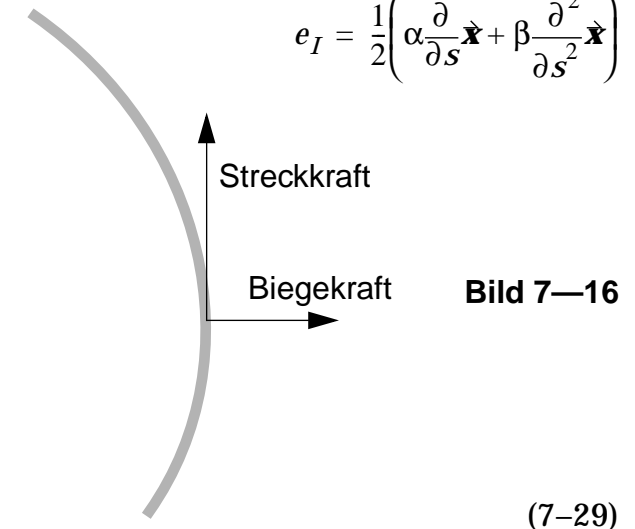
Als äußere Energiedichte  $e_A$  wirkt die Lageenergie aufgrund der durch den lokalen Gradienten der Bildsignalfunktion  $g$  gegebenen Topologie, Gl. (7-29).  $\nabla g(\vec{x})$  entspricht darin der Geländetiefe, Bild 7—17.

Die Kontur verlagert sich aufgrund des äußeren Kraftfeldes so, dass ein 'Hineingleiten' in Täler oder Gräben die Energie für ihre Verformung liefert.

1. Blake A, Isard M: Active Contours, Springer London 1998

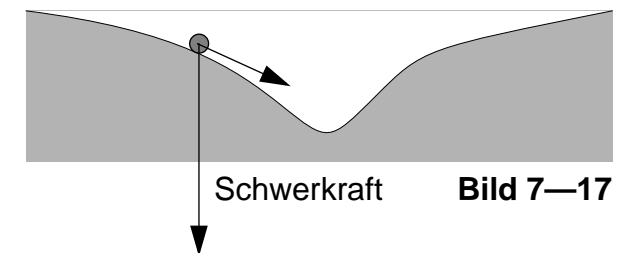
(7-28)

$$e_I = \frac{1}{2} \left( \alpha \frac{\partial}{\partial s} \vec{x} + \beta \frac{\partial^2}{\partial s^2} \vec{x} \right)$$



(7-29)

$$e_A = -|\nabla g(\vec{x})|$$



Der Modellvorgang der Konturanpassung wird also zu einer Bestimmung des Gleichgewichts zwischen Verformungs- und Lageenergie: Gesucht wird das Minimum der Gesamtenergie über die Konturlänge  $S$ , Gl. (7-30).

$$E_{\text{ges}} = \int_S (e_I + e_a) ds \quad (7-30)$$

Für die Bestimmung des Minimums der Konturenergie wird die Euler-Lagrange-Gleichung (7-31) herangezogen, deren Lösung das gesuchte Minimum ergibt. Die Bearbeitung erfordert eine diskrete Approximation der Konturlinie. Lösungsansätze wurden mit finiten Differenzen oder FEM-Modellierungen (Abschnitt 7.2.4) erfolgreich bearbeitet.

$$-\alpha \frac{\partial^2 \mathbf{x}}{\partial s^2} + \beta \frac{\partial^4 \mathbf{x}}{\partial s^4} = -\nabla e_A(\mathbf{x}(s)) \quad (7-31)$$

### 7.3.3 Morphologische Operationen auf binären Bildern

Ein anderes Modell legt das Analogon modellierbarer Körper zugrunde, die mit strukturbildenden Werkzeugen bearbeitet werden: Der Form und Bewegung des Werkzeugs folgend kann Material von der Oberfläche abgetragen, 'erodiert' oder der Körper durch Auftragen von Material ausgedehnt, 'dilatiert' werden. Im Zweidimensionalen führt dieses Modell zu Operationen, die auf Regionendarstellungen aufsetzen, im einfachsten Fall auf binären Darstellungen kompakter Objekte  $g_O = 1$  vor einem Hintergrund  $g_H = 0$ . Mit dem strukturgebenden Werkzeug wird die Gestalt beeinflusst. Deshalb wurde für diese Klassen von Operationen der Terminus 'morphologische Operationen' geprägt.

Morphologische Operationen auf binären Bildern können als Mengenoperationen formuliert werden. Für das Verändern der Gestalt eines binären Objektes — Pixelmenge  $X$  — wird ein sogenanntes Strukturelement — ebenfalls ein binäres Objekt, Pixelmenge  $S$  — mit einem Ankerpunkt  $x_S$  definiert. Beispiele für Strukturelemente sind Linie, Kreuz, Quadrat oder Kreisscheibe, Bild 7—18.

Basisoperationen sind die *Erosion* und die *Dilatation*. Das Strukturelement wird dabei an seinem Ankerpunkt zeilen- und spaltenweise über alle Pixel des Bildes geführt:

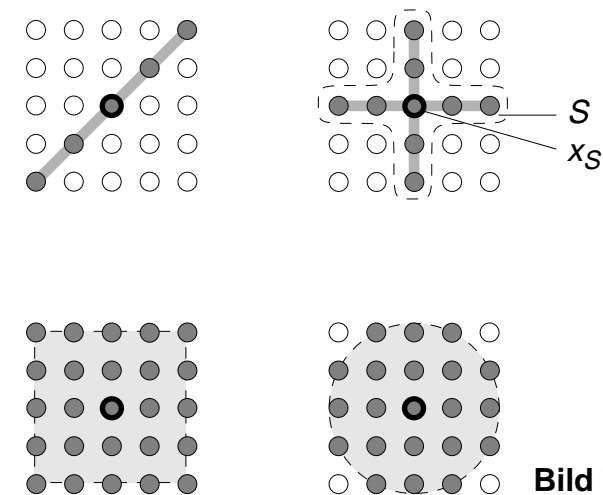


Bild 7—18

- Bei der Erosion  $\varepsilon_S$  wird an jeder Position gefragt: „Passt das Strukturelement vollständig in die Menge  $X$ ?“. Im negativen Fall wird das Ankerpixel auf  $\emptyset$  gesetzt. Das Verfahren kann als Mengenaufgabe nach Gl. (7-32) formuliert werden.
- Bei der Dilatation  $\delta_S$  wird komplementär gefragt: „Berührt das Strukturelement die Menge  $X$ ?“. Im positiven Fall wird das Ankerpixel auf 1 gesetzt. Entsprechend wird daraus die Mengenaufgabe Gl. (7-33).

Als alternative Formulierungen können die *Minkowski-Subtraktion* bzw. die *Minkowski-Addition* dienen, wenn das Strukturelement jeweils um  $180^\circ$  rotiert wird.

Erosion und Dilatation verändern nicht nur die Gestalt der Kontur eines Objektes, sie verändern auch dessen Ausdehnung. Letzteres kann unerwünscht sein. Abhilfe schaffen Operationen, die aus aufeinanderfolgender Erosion und Dilation zusammengesetzt sind:

- Die *morphologische Öffnung* ('opening') ist die Aufeinanderfolge einer Erosion und einer Dilatation mit dem um  $180^\circ$  rotierten Strukturelement der Erosion. Sie ergibt nach Gl. (7-34) die Vereinigung aller Strukturelemente, die in  $X$  hineinpassen.
- Die *morphologische Schließung* ('closing') ist die Aufeinanderfolge einer Dilatation und einer Erosion mit dem um  $180^\circ$  rotierten Strukturelement der Dilatation. Hier ist eine komplementäre Betrachtung am einfachsten: Der Hintergrund ergibt sich nach Gl. (7-35) aus der Vereinigung aller Strukturelemente, die  $X$  nicht schneiden.

Öffnung und Schließung mit einem flächigen Strukturelement sind vor allem für die Überarbeitung einer ersten Segmentierung gut geeignet. Mit einer Öffnung werden Brücken zwischen Regionen aufgelöst, kleine Störobjekte ausgelöscht und Ausstülpungen der Konturen abgeschliffen. Eine Schließung verbindet Regionen, die nur durch schmale Kanäle getrennt sind, schließt Löcher in Regionen und füllt Einbuchtungen der Konturen auf. Überwiegend glatte Konturabschnitte werden nicht verändert.

$$\varepsilon_S(X) = \{x_S \mid S \subseteq X\} \quad (7-32)$$

$$\delta_S(X) = \{x_S \mid S \cap X \neq \emptyset\} \quad (7-33)$$

$$\gamma_S(X) = x_S \cup \{S \mid S \subseteq X\} \quad (7-34)$$

$$\eta_S(X) = x_S \cap \{S^c \mid X \subseteq S^c\} \quad (7-35)$$



Die Implementierung morphologischer Operationen vereinfacht sich durch Anwendung einiger grundlegender Rechenregeln. So kann ein komplizierteres Strukturelement mittels Dilatation eines ersten Strukturelementes mit einem zweiten Strukturelement erzeugt werden (z.B. ein Rechteck aus zwei diagonalen Linienelementen, ein Sechseck als Näherung für einen Kreis aus drei um jeweils  $60^\circ$  versetzten Linienelementen). Einer Erosion oder Dilatation mit dem so entstandenem Strukturelement entspricht die aufeinanderfolgende Erosion bzw. Dilatation mit dessen Komponenten. Weiterhin gilt, dass einer Erosion die Dilatation auf dem invertierten Bild entspricht.

Ein weiterer Typ der morphologischen Operationen ist die *Skelettierung*, Bild 7—19. Die Skelettierung dient dem Auffinden einer Art Mittellinie (vergleichbar dem Leitgefäßgerüst eines Blattes) bei einem komplexer gestalteten Objekt, um daran Gestaltmerkmale festzulegen. Eine Fülle von Verfahrensvarianten liegt hier vor, von denen eine Auswahl wiedergegeben wird:

- Das Verfahren der *'maximalen Scheiben'* beruht auf der Bestimmung der Mittelpunkte aller vollständig von der Kontur umschlossenen kreisförmigen Scheiben, die die Kontur an mindestens zwei unterschiedlichen Punkten tangieren.
- Beim Verfahren der *'minimale Wege'* ist das Skelett durch alle Pixel gegeben, von denen gleichlange kürzeste Wege zu zwei unterschiedlichen Punkten auf der Kontur gefunden werden können.
- Bei der *'Grasfeuerausbreitung'* wird das Objekt jeweils solange erodiert, bis sich zwei von unterschiedlichen Konturabschnitten aufeinander zu bewegende 'Feuerfronten' treffen. Diese Punkte werden als Skelettpixel markiert.
- Unter Anwendung der *Distanzfunktion* (siehe Seite 176) wird für jedes Pixel innerhalb der Kontur, von der Kontur aus beginnend, die Distanz zur Kontur bestimmt, Bild 7—20. Nimmt die Distanz wieder ab, war ein Skelettpixel erreicht, das mit dem Distanzmaß etikettiert wird. Dies ist das einzige unter den genannten Verfahren, das eine Rekonstruktion der Kontur aus dem Skelett zulässt.

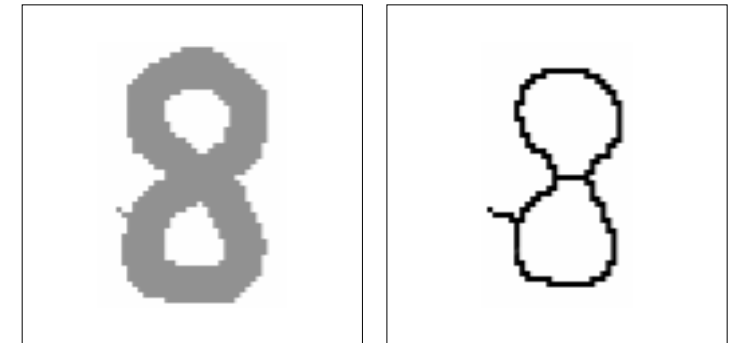


Bild 7—19

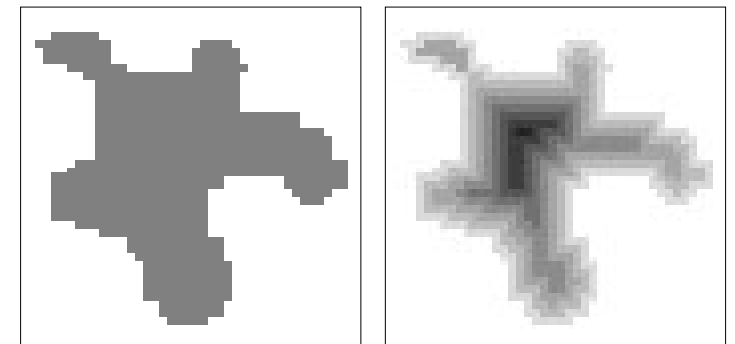


Bild 7—20



## 8 Klassifikation

*Präzision ist der Feind der Relevanz; alles ist „a matter of degree“, eine Frage des Grades*  
— Lofti Zadeh, Pionier der Fuzzy-Logik

Wir nehmen eine Szene wahr und fassen das Gesehene in Worte, um uns darüber zu verständigen, wir ziehen aus dem Gesehenen Folgerungen, die unser Handeln leiten. Erkennen und Verstehen, das ist ein Abstraktionsprozess, mit dem wir die Vielfalt der wahrgenommenen Objekte in ein Ordnungsschema einordnen. Dieser Ordnungsprozess setzt nun aber voraus, dass wir in der Lage sind, Objekte trotz mannigfaltigen Variationen ihrer Erscheinung “in die richtige Schublade zu stecken”, d.h. sie mit den korrekten Sammelbegriffen zu belegen. Wie gehen wir dabei vor?

Ein Vergleich auf “Passgenauigkeit” zwischen der Abbildung eines Objektes und einer bildlichen Vorlage wird kaum Aussicht auf Erfolg haben. Im allgemeinen existiert eine solche idealtypische Bildvorlage auch gar nicht. Eine brauchbarere Hypothese geht davon aus, dass wir aus dem Abbild des Objektes *Merkmale* destillieren, die für die Einordnung in das vorgegebene Ordnungsschema relevant sind. Ordnungsschemata sind wiederum vom Kontext unserer Situation, Umgebung, Aufgabe abhängig. Entsprechend variieren auch Relevanz und Bedeutung der Merkmale. Im Möbelhaus z.B. ordnen wir Gegenstände mit einer horizontalen Platte und vier etwa vertikal angebrachten Stäben in die Kategorie “Tische” ein. Wir sind aber auch in der Lage, erhebliche Abweichungen zu tolerieren, etwa einen Tisch mit nur einem Mittelfuss oder einen Frisiertisch trotz der aufgesetzten Spiegel zu akzeptieren, einen Stuhl dagegen (horizontale Sitzfläche ebenfalls mit vier Beinen) aber zu verwerfen. Die Unterschiede sind kaum scharf zu definieren und dennoch fällen wir unausgesetzt unsere Entscheidungen — meistens unbewusst und meistens korrekt!

Entscheidungsprozesse im Sinne eines Einordnens sind die Grundlage aller Schritte zur Informationsverdichtung. Wir treffen sie auf allen Ebenen innerhalb des Bildanalyseprozesses: Auf der Pixel-Ebene wird entschieden,

## 8 Klassifikation

ob Pixel aufgrund von Homogenitätskriterien mit ihrer Nachbarschaft zu verschmelzen sind (eine Ja/Nein-Entscheidung), auf der Gestalt-Ebene wird entschieden, ob Segmente nach einem Verschmelzen oder Teilen Formkriterien erwarteter Objekte besser erfüllen, auf der Objektebene wird schließlich entschieden, wie die Objekte eines Bildes begrifflich einzuordnen sind. Als Voraussetzung muss jeweils eine zweckmäßige Einteilung der Menge der Möglichkeiten in eine zumeist kleine Zahl von Kategorien (im Minimalfall zwei) gefunden werden. Solche Kategorien ("Schubladen") werden als *Klassen* bezeichnet. Der Entscheidungsprozess, der zu der angestrebten Einordnung führt, wird *Klassifikation* genannt. Objekte, die in eine bestimmte Klasse eingeordnet worden sind, sind *Instanzen* dieser Klassen.

Auch mit der Klassifikation treffen wir wiederum auf einen Aufgabentyp, den wir alltäglich in mannigfaltiger Gestalt bewältigen. Der dazu erforderlichen intellektuellen Leistung werden wir uns kaum bewusst. So neigen wir auch hier wieder dazu, die Schwierigkeiten für maschinelle Lösungen weit zu unterschätzen, denn gerade in der Bildanalyse treffen wir auf einige besondere Schwierigkeiten der Klassifikation:

- Aufgrund der Ganzheitlichkeit der Repräsentation eines Objektes oder einer Szene durch das Bild (Bilder sind nicht wie eine textliche Repräsentation in semantische Einheiten zu zerlegen) müssen aus Bildern zunächst Merkmale extrahiert werden, die bezüglich eines vorliegenden Klassensystems relevant sind.
- Einfache kausale Beziehungen zwischen dem Klassensystem und Merkmalen, die sich unmittelbar aus den Bilddaten ableiten lassen, sind — insbesondere auf den höheren Ebenen der Bildanalyse — im allgemeinen nicht gegeben.
- Den zur Klassifikation geeigneten Merkmalen sind systematische und stochastische Störungen überlagert.
- Es fehlt darüber hinaus bereits das a-priori-Wissen, um ein Klassensystem oder idealtypische Instanzen (*Prototypen*) vorzugeben.

Angesichts der Vielfalt unterschiedlicher Teilaufgaben und vorgegebener Randbedingungen ist eine Fülle unterschiedlicher Strategien und Verfahren zur Klassifikation entwickelt worden. Im folgenden soll versucht werden, einige der wichtigsten dieser Ansätze auf der Basis einiger grundlegender Überlegungen zum Thema Klassifikation zu beschreiben.

### 8.1 Der Merkmalsraum

Ein Objekt sei durch  $d$  Merkmale beschrieben. Dann können diese zu einem  $d$ -dimensionalen *Merkmalsvektor*  $\mathbf{m}$ , Gl. (8-1), zusammengefasst und in einem  $d$ -dimensionalen Merkmalsraum, Bild 8-1, dargestellt werden. Für den Merkmalsvektor ist auch der Begriff *Muster* geläufig, den wir hier jedoch aus Gründen der semantischen Mehrdeutigkeit vermeiden wollen.

$$\mathbf{m} = \begin{Bmatrix} m_1 \\ m_2 \\ \dots \\ m_d \end{Bmatrix} \quad (8-1)$$

Klassen zu definieren heißt nun, den Merkmalsraum in disjunkte Teilräume aufzuteilen, die jeweils die Merkmalsvektoren von Instanzen derselben Klasse ausschließlich und komplett enthalten. Die Aufgabe der Klassifikation wird es damit, eine *optimale* Partition des Merkmalsraumes zu finden. Dabei bedeutet "optimal", dass bei einem Test mit den Merkmalsvektoren beliebiger Objekte möglichst wenige Fehlklassifikationen geleistet werden. So wird eine Fehlklassifikation dann erfolgen, wenn der Merkmalsvektor infolge systematischer oder stochastischer Variationen einzelner Merkmale in eine nicht zutreffende Partition des Merkmalsraumes fällt, d.h. wenn das Objekt bezüglich der zur Klassifikation herangezogenen Merkmale Instanzen einer Nachbarklasse zu ähnlich ist.

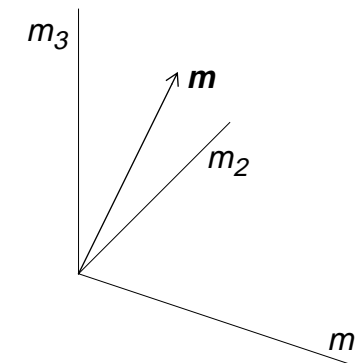


Bild 8-1

Für die Klassifikation können unterschiedliche Voraussetzungen bestehen:

1. Prototypen aller Klassen existieren und von allen Prototypen liegen genaue Beschreibungen vor.
2. Prototypen liegen *nicht* vor, aber für alle Klassen sind Instanzen in größerer Zahl gegeben, so dass ein Training des Klassifikators möglich ist. Dieser Vorgang wird *überwachtes Lernen* genannt.

3. A priori klassifizierte Merkmalsvektoren liegen *nicht* vor. Der Klassifikator muss in diesem Fall eine sinnvolle Partition des Merkmalsraumes allein aus den vorliegenden Merkmalsvektoren erstellen. Dies wird als *unüberwachtes Lernen* bezeichnet.

## 8.2 Eine Metrik für die Ähnlichkeit

Der Ausgangspunkt jeder Klassifikation ist die Hypothese, dass Instanzen derselben Klasse hinsichtlich der ausgewählten Merkmale einander ähnlicher sind, als Instanzen unterschiedlicher Klassen. Nachdem bereits die Darstellung der zur Unterscheidung herangezogenen Objekteigenschaften in einem Merkmalsraum eine Quantifizierung impliziert, soll für die Bemessung einer Ähnlichkeit oder Unähnlichkeit zweier Vektoren im Merkmalsraum ein quantifizierbares Maß gefunden werden. Dem muss zunächst eine Metrik zugrunde gelegt werden. Sehr gebräuchlich sind die sogenannten  $L_p$ -Metriken nach *Minkowski*, Gl. (8-2). Geläufig sind vor allem die  $L_p$ -Metriken mit den Exponenten 1 und 2:

- $p = 1$  als *Schachbrett-* oder *Manhattan-*Abstand, engl. “city block metric” (identisch mit der *Distanz* in einem einfachen ebenen Graphen aus orthogonalen, einheitlich langen Kanten, siehe Seite 177)
- $p = 2$  als *euklidischer* Abstand.

Abhängig von  $p$  führt die Richtung des Abstandsvektors im Merkmalsraum (Differenz zwischen zwei Merkmalsvektoren) zu unterschiedlichen Gewichtungen, siehe Bild 8—2. Nur der am häufigsten verwendete euklidische Abstand ist isotrop. Vergleichende Untersuchungen haben aber gezeigt, dass der bezüglich des Rechenaufwands anspruchslosere Manhattan-Abstand kaum signifikant unterschiedliche Ergebnisse erbringt.<sup>1</sup>

$$L_p(\mathbf{m}, \mathbf{m}') = \sqrt[p]{\sum_d |m_d - m'_d|^p} \quad (8-2)$$

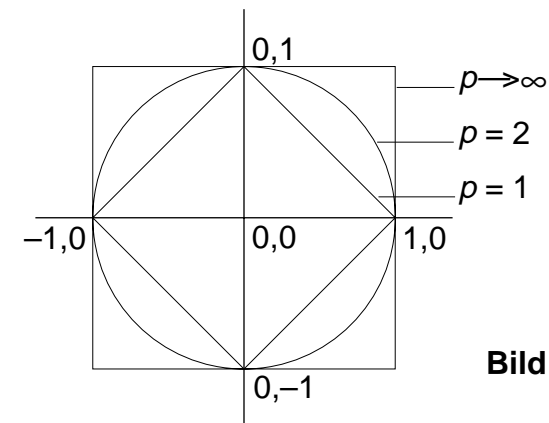


Bild 8—2

1. Salzberg S: Distance Metrics for Instance-based Learning; in: Methodologies for Intelligent Systems (ISMIS '91), Lecture Notes in Computer Science 542, 399—408, 1991

### 8.3 Das Entscheidungsproblem

Unabhängig davon, welche der drei Ausgangssituationen gegeben ist, bestimmt das Ähnlichkeitsmaß, d. h. der Abstand im Merkmalsraum, die Entscheidung über die Zugehörigkeit eines Objektes zu einer Klasse. Die generelle Schwierigkeit einer Klassifikation liegt nun darin, dass die Menge der Instanzen, die einer bestimmten Klasse zuzuordnen sind, eine Verteilung darstellt, die über einen Bereich des Merkmalsraumes ausgedehnt ist. Die Schwerpunkte solcher Raumbereiche können im allgemeinen als *Erwartung* einer unbekanntem Instanz der jeweiligen Klasse betrachtet werden. Die Grenzflächen der Bereiche können aber nur unscharf bestimmt werden, denn Abweichungen des Merkmalsvektors individueller Instanzen von der Erwartung sind vor allem durch stochastische Einflüsse bedingt.

Soll nun zwischen zwei Klassenschwerpunkten eine Entscheidungsfläche eingezeichnet werden, stellt sich die Aufgabe, deren Lage so zu bestimmen, dass *die Zahl der Fehlzuordnungen minimiert* wird. Die einfachste Situation liegt vor, wenn a priori

1. die Zahl  $I$  der Klassen  $K_1 \dots K_I$ ,
2. die Schwerpunkte  $\hat{\mathbf{m}}_1 \dots \hat{\mathbf{m}}_I$  aller Klassen<sup>1</sup> und
3. die Verteilungen innerhalb der Klassen als parametrisierte Funktionen bekannt sind. Betrachten wir die Merkmalsvektoren der Instanzen aller Klassen als Resultat von Zufallsprozessen, ist jeder Merkmalsvektor die Realisation einer multivariaten Zufallsvariablen. Die Wahrscheinlichkeitsfunktion  $p(\mathbf{m}/K_j)$  für das Auftreten einer Realisation  $\mathbf{m}$  als Instanz einer Klasse  $K_j$  stellt dann eine Verteilungsdichte im Merkmalsraum dar.

Häufig können die wahren Verteilungsdichten durch multivariate Gauß-Verteilungen Gl. (8-3) approximiert werden, deren Parameter  $\hat{\mathbf{m}}_j$ ,  $(\mathbf{M}_j)$ ,  $i = 1 \dots I$  aus den Trainingsdaten geschätzt werden. Darin ist  $(\mathbf{M}_j)$  die Kovarianzmatrix der Klasse  $K_j$  nach Gl. (8-4).

1.  $\hat{\mathbf{m}}_j = E\{\mathbf{m}(K_j)\}$  ist die Erwartung für Merkmalsvektoren von Instanzen der Klasse  $K_j$

$$p(\mathbf{m}|K_j) = \frac{1}{\sqrt{(2\pi)^d \det(\mathbf{M}_j)}} \cdot e^{-\frac{1}{2} \langle \mathbf{m} - \hat{\mathbf{m}}_j \rangle^T (\mathbf{M}_j)^{-1} \langle \mathbf{m} - \hat{\mathbf{m}}_j \rangle} \quad (8-3)$$

$$(\mathbf{M}_j) = E\{\langle \mathbf{m} - \hat{\mathbf{m}}_j \rangle \langle \mathbf{m} - \hat{\mathbf{m}}_j \rangle^T\} \quad (8-4)$$

## 8.3.1 Die Bayes-Entscheidungsregel

Eine Wahrscheinlichkeits-theoretisch begründete Regel, die für die Entscheidung der Klassenzugehörigkeit neuer Objekte herangezogen werden kann, ist die *Bayes-Entscheidungsregel*  $r(\mathbf{m})$ :

Klassifiziere  $\mathbf{m}$  in diejenige Klasse  $K_i$  für die die *a-posteriori*-Wahrscheinlichkeit  $p(K_i|\mathbf{m})$  am größten ist, Gl. (8-5)!

Die *a-posteriori*-Wahrscheinlichkeit berücksichtigt Vorwissen über die Besetzungstärken aller Klassen. Nach dem Satz von *Bayes* berechnet sie sich als Produkt aus der so genannten klassenbezogenen Wahrscheinlichkeit  $p(\mathbf{m}|K_j)$  für eine Beobachtung  $\mathbf{m}$  gegeben eine Klasse  $K_j$  und der *a-priori*-Wahrscheinlichkeit  $p(K_j)$  der einzelnen Klassen, Gl. (8-6). Der Nenner  $p(\mathbf{m})$  beschreibt nach Gl. (8-7) die Grundwahrscheinlichkeit für das Auftreten der Realisation, unabhängig von ihrer Klassenzuordnung. Er dient also lediglich der Normierung und kann für die Klassifikation vernachlässigt werden.

Man kann zeigen, dass die Zahl der Fehlklassifikationen minimiert wird, wenn das Verhältnis der *a-posteriori*-Wahrscheinlichkeiten der jeweils in Betracht gezogenen Klasse  $K_i$  und der übrigen Klassen  $K_j$  für alle  $j \neq i$  größer als 1 ist, Gl. (8-8). Instanzen  $\mathbf{m}_E$  mit gleichgroßen *a-posteriori*-Wahrscheinlichkeiten definieren somit eine Trenn- oder Entscheidungsfläche. Sie kann in dieser Form implizit beschrieben werden. Zweckmäßig ist eine logarithmische Darstellung, Gl. (8-9). Besitzt der Merkmalsraum nur eine einzige Dimension, degeneriert die Entscheidungsfläche zu einem Schwellwert  $m_E$ .

Die Klassifizierungsaufgabe kann so gelagert sein, dass Fehlklassifikationen abhängig von der Klassenzugehörigkeit einer Instanz unterschiedliche Folgen haben (so kann zum Beispiel in der medizinischen Diagnostik die Nichtentdeckung eines Befundes weit schwerer wiegende Folgen haben als ein scheinbarer Befund). In diesem Fall müssen Kostenterme  $c_{i,j}$  in die Berechnung der Trennfläche einbezogen werden, Gl. (8-10), da die Bayes-Regel weder die absolute noch die relative Zahl der Fehlentscheidungen angibt.

$$\mathbf{m} \rightarrow r(\mathbf{m}) = \operatorname{argmax}\{p(K_i|\mathbf{m})\} \quad (8-5)$$

$$p(K_i|\mathbf{m}) = \frac{p(\mathbf{m}|K_i) \cdot p(K_i)}{p(\mathbf{m})} \quad (8-6)$$

$$p(\mathbf{m}) = \sum_i p(\mathbf{m}|K_i) \cdot p(K_i) \quad (8-7)$$

$$\frac{p(K_i|\mathbf{m})}{p(K_j|\mathbf{m})} = \frac{p(\mathbf{m}|K_i)}{p(\mathbf{m}|K_j)} \cdot \frac{p(K_i)}{p(K_j)} > 1 \quad (8-8)$$

$$\frac{p(\mathbf{m}_E|K_i)}{p(\mathbf{m}_E|K_j)} \cdot \frac{p(K_i)}{p(K_j)} = 1$$

$$\ln \frac{p(\mathbf{m}_E|K_i)}{p(\mathbf{m}_E|K_j)} = \ln \frac{p(K_i)}{p(K_j)} \quad (8-9)$$

$$\frac{p(\mathbf{m}_E|K_i)}{p(\mathbf{m}_E|K_j)} \cdot \frac{c_i \cdot p(K_i)}{c_j \cdot p(K_j)} = 1 \quad (8-10)$$



8.3.2 Ein einfaches Beispiel

Dies sei am Beispiel einer 2-Klassen-Entscheidung in einem eindimensionalen Merkmalsraum erläutert. Wir greifen dazu auf die Detektionsaufgabe auf Seite 25 zurück. Dort waren die  $10^4$  Pixel einer Sensormatrix in Instanzen der Klassen  $K_1 = \text{“Ereignis”}$  (erwartete Anzahl 5) und  $K_2 = \text{“Hintergrund”}$  zu klassifizieren. Vereinfachend ist den Instanzen beider Klassen ein gaußscher Rauschprozess mit identischer Varianz  $\sigma^2$  überlagert, so dass sich Wahrscheinlichkeiten gemäß Gl. (8-11) ergeben.

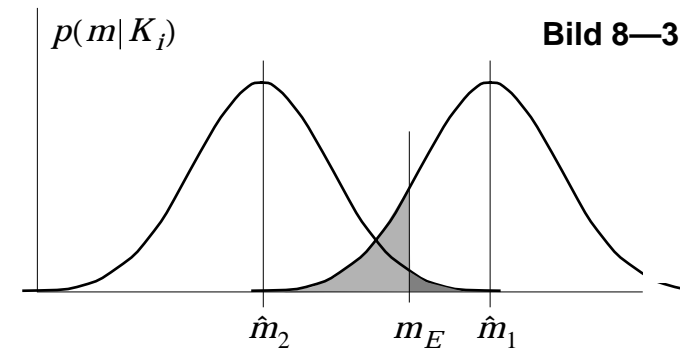
$$p(\mathbf{m} | K_{1,2}) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(m - \hat{m}_{1,2})^2}{2\sigma^2}} \quad (8-11)$$

$$\ln \frac{e^{-(m_E - \hat{m}_1)^2/2\sigma^2}}{e^{-(m_E - \hat{m}_2)^2/2\sigma^2}} = \ln \frac{10^4}{5}$$

$$\frac{(m_E - \hat{m}_2)^2}{2\sigma^2} - \frac{(m_E - \hat{m}_1)^2}{2\sigma^2} = 7,6$$

$$m_E = \frac{\hat{m}_1 + \hat{m}_2}{2} + \frac{7,6 \cdot \sigma^2}{\hat{m}_1 - \hat{m}_2} \quad (8-12)$$

Eingesetzt in Gl. (8-9) wird die Entscheidungsschwelle  $m_E$  umso mehr aus der Abstandsmitte zwischen den Klassenerwartungswerten  $\hat{m}_1$  und  $\hat{m}_2$  verschoben, je unterschiedlicher die a-priori-Wahrscheinlichkeiten der beiden Klassen sind, Gl. (8-12). So würde im oben zitierten Beispiel bei einem Abstand der Klassenerwartungswerte von  $5\sigma$  die Entscheidungsschwelle  $m_E$  auf näherungsweise  $4\sigma$  von  $\hat{m}_2$  abrücken. Die diesseits und jenseits der Schwelle liegenden Restflächen unter den Verteilungen (in Bild 8-3 nicht maßstabsgerecht dargestellt) sind ein Maß für die Fehlentscheidungsrate. Gewichtet mit den Klassenwahrscheinlichkeiten  $p(K_j)$  repräsentieren sie die (minimierte) Gesamtzahl der Fehlentscheidungen — in diesem Fall ca. 0,8 falsch-negative und 0,3 falsch-positive Entscheidungen pro Experiment.



8.3.3 Entscheidungsflächen im mehrdimensionalen Merkmalsraum

Setzen wir weiterhin Normalverteilungen nach Gl. (8-3) auf Seite 199 für die Klassenwahrscheinlichkeitsdichten  $p(\mathbf{m} | K_j)$  voraus, dann ergeben sich als Flächen gleicher Dichte Hyperellipsoide mit der Gleichung (8-13). Letztere definiert den Mahalanobis-Abstand<sup>1</sup> zum Schwerpunkt  $\hat{m}_i$  der Klasse.

$$\langle \mathbf{m} - \hat{\mathbf{m}}_i \rangle^T (\mathbf{M}_i)^{-1} \langle \mathbf{m} - \hat{\mathbf{m}}_i \rangle = const \quad (8-13)$$

1. Der Mahalanobis-Abstand ist ein Ähnlichkeitsmaß, das unterschiedliche Ausdehnungen einer Klasse in den Richtungen des Merkmalsraumes berücksichtigt.

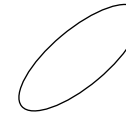
Ein Klassifikator, der Merkmalsvektoren  $I$  Klassen zuordnen soll, muss entsprechend aus  $I$  Diskriminanzfunktionen bestehen. Jeder zu klassifizierende Merkmalsvektor wird nun in alle Diskriminanzfunktionen eingesetzt. Es wird derjenigen Klasse, dessen Diskriminanzfunktion den größten Wert ergibt, zugeordnet. Entscheidungsflächen, d.h. Trennflächen zwischen den Klassen im Merkmalsraum, sind damit durch die Menge aller Punkte definiert, an denen die Diskriminanzfunktion mindestens zweier Klassen denselben Wert besitzt.

Als Diskriminanzfunktion lassen sich die a-posteriori-Wahrscheinlichkeiten  $p(K_i/\mathbf{m})$  nach Gl. (8–6) auf Seite 200 verwenden. Setzen wir in diese multivariate Normalverteilungen  $p(\mathbf{m}/K_j)$  ein, dann ergeben sich quadratische Diskriminanzfunktionen. Die Entscheidungsflächen werden somit quadratische Flächen — Hyperkugeln, Hyperellipsoide, Hyperparaboloide, Hyperhyperboloide. Die Besetzung der Kovarianzmatrix gibt Aufschluß über die Orientierung dieser Flächen, wie Bild 8–4 am Beispiel elliptischer Trennlinien im Fall eines zweidimensionalen Merkmalsraum zeigt. Dies weist darauf hin, dass das Entscheidungsproblem gegebenenfalls mit geeigneten Koordinatentransformationen vereinfacht werden kann.

Als Sonderfall ergeben sich *Hyperebenen*, wenn alle Merkmale voneinander statistisch unabhängig sind und eine identische Varianz  $\sigma^2$  aufweisen. Die Entscheidungs-Hyperebene zwischen zwei Klassen verläuft dann senkrecht zur Verbindungslinie zwischen den Klassenschwerpunkten durch einen Punkt auf dieser Linie, der analog zu dem eindimensionalen Fall im vorangehenden Beispiel zu berechnen ist. Sind sogar alle a-priori-Wahrscheinlichkeiten gleich, dann sind die Entscheidungsflächen die Mittelsenkrechten auf den Verbindungslinien zwischen jeweils zwei der Klassenschwerpunkte. Ein solcher Klassifikator wird als *Minimaler-Abstand-Klassifikator* ('minimum distance classifier') bezeichnet, da er einen Merkmalsvektor derjenigen Klasse zuordnet, deren Schwerpunkt ihm am nächsten liegt.

Kovarianzmatrix beliebig:

beliebig orientierte Ellipse



Kovarianzmatrix diagonal:

Hauptachsen parallel zu den Koordinaten



Kovarianzmatrix = Identitätsmatrix:

Isotropie (Kreis)

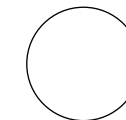


Bild 8—4

### 8.4 Überwacht lernende Klassifikatoren

Wir sind zunächst davon ausgegangen, dass für alle Klassen Prototypen vorliegen und dass die Merkmalsvektoren der Objekte als gestörte Realisation der Prototypen interpretiert werden können. In der Praxis muss das a-priori-Wissen über die Klassenerwartungswerte und die überlagerten Störprozesse jedoch im allgemeinen erst experimentell gewonnen werden. Die Verfahren, mit denen Klassifizierer dieses Wissen aus den Eigenschaften einer Trainingsmenge von Instanzen mit bekannter Klassenzugehörigkeit “erlernen”, werden in *parametrische* und *nicht-parametrische* Ansätze eingeteilt. Das Attribut *überwacht-lernend* unterscheidet sie von den *unüberwacht lernenden* Verfahren, Abschnitt 9.5, die für Situationen entworfen wurden, in denen eine Trainingsmenge nicht angeboten werden kann.

#### 8.4.1 Parametrische Klassifizierer

Parametrische Klassifikationsverfahren<sup>1</sup> setzen voraus, dass die den Instanzen aller Klassen überlagerten Störprozesse als Zufallsprozesse mit einer parametrisierbaren Verteilungsdichtefunktion — gewöhnlich einer Normalverteilung,  $p(m|K_j)$  — gehorchen, deren Parameter zunächst unbekannt sind. Liegt eine genügend große Menge von Instanzen bekannter Klassenzugehörigkeit vor, dann können die Parameter aus deren Merkmalsvektoren geschätzt werden. Als Algorithmus wird hierfür der so genannte Maximum-Likelihood-Schätzer eingesetzt. Werden multivariate Normalverteilungen vorgegeben, dann errechnen sich deren Parameter — Mittelwert bzw. Schwerpunkt  $\hat{\mathbf{m}}_i$  und Kovarianzmatrix  $(\mathbf{M}_i)$  — nach Gln. (8-14) und (8-15).

Eine Möglichkeit zur Vereinfachung des Klassifikationsproblems bietet sich in der linearen Diskriminanzanalyse. Ihr Ziel ist eine Transformation des Merkmalsraumes, die zu einer geringeren Zahl von Merkmalen (Reduktion

$$\hat{\mathbf{m}}_i = E\{\mathbf{m}_{n,i}\} = \frac{1}{N_i} \sum_{n=1}^{N_i} \mathbf{m}_{n,i} \quad (8-14)$$

$$(\mathbf{M}_i) = \frac{1}{N_i} \sum_{n=1}^{N_i} \langle \mathbf{m}_{n,i} - \hat{\mathbf{m}}_i \rangle \langle \mathbf{m}_{n,i} - \hat{\mathbf{m}}_i \rangle^T \quad (8-15)$$

1. Duda RO, Hart PE: Pattern Classification and Scene Analyses; Wiley, New York 1973

der Dimension) führt und dabei die Merkmalsvektoren so anordnet, dass parametrische Verfahren angewandt werden können. Die Achsen des transformierten Merkmalsraumes entsprechen neuen Merkmalen, die als Linearkombination aus den ursprünglichen Merkmalen hervorgehen. Die Zahl der transformierten Merkmale, d.h. die Dimension des Vektors  $\mathbf{m}$  wird dabei auf  $I-1$  reduziert, da  $I$  Klassenschwerpunkte maximal einen  $(I-1)$ -dimensionalen Unterraum aufspannen können. Die Transformationsmatrix zwischen den ursprünglichen und den neuen Merkmalen wird nun so bestimmt, dass das Verhältnis der Varianz zwischen den Klassen zur Varianz innerhalb der Klassen maximiert wird.

Die lineare Diskriminanzanalyse ist mit der *Hauptachsen-Transformation* ('principal component analysis'), auch *Karhunen-Loève-Transformation* (KLT) verwandt. Während jedoch die lineare Diskriminanzanalyse die Merkmale nach der besten Klassentrennung optimiert, richtet die Hauptachsen-Transformation die Merkmale nach der größten Varianz über der gesamten Trainingsmenge aus. Der Unterschied soll an einem einfachen Zwei-Klassen-Problem erläutert werden, Bild 8—5. Hierfür existiert nur ein einziges optimales Merkmal. Die lineare Diskriminanzanalyse ermittelt in diesem Fall die durchgezogene Linie als eine (einzige) Merkmalsdimension. Die Hauptachsen-Transformation ermittelt lediglich die Richtung der größten Varianz *aller* Merkmalsvektoren (punktierte Linie). Sie trägt somit nicht (direkt) zur Klassentrennung bei.

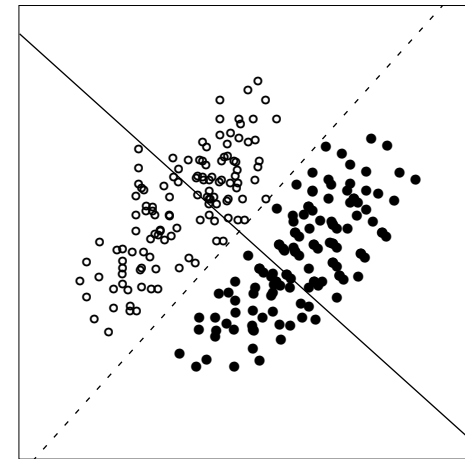


Bild 8—5

#### 8.4.2 Nichtparametrische Klassifikationsverfahren

Oft ist es nicht möglich, mit wenigen Parametern beschreibbare Funktionale für die Klassenwahrscheinlichkeitsdichte  $p(\mathbf{m}/K_j)$  der Trainingsmenge anzugeben. Auch die Wahrscheinlichkeiten der Klassen  $p(K_j)$  sind meist unbekannt. In diesen Fällen muss sich der Klassifikator allein auf die Repräsentativität der Trainings-Merkmalsvektoren stützen. Grundsätzlich unterschiedliche Wege gehen Verfahren, die die lokale Dichte der Merkmals-

vektoren für die Klassenzugehörigkeits-Entscheidung zugrunde legen, gegenüber Verfahren, die Diskriminanzfunktionen eines vorgegebenen Typs (z. B. stückweise lineare Entscheidungshyperebenen) in den Merkmalsraum einpassen.

Zum ersten Verfahrenstyp gehört der Nächste-Nachbar-Klassifikator.<sup>1</sup> Im einfachsten Fall, der 1-NN-Regel, wird ein Merkmalsvektor in diejenige Klasse eingeordnet, der sein nächster Nachbar im Merkmalsraum angehört. Als Ähnlichkeitsmaß wird im allgemeinen der euklidische Abstand gewählt (siehe Seite 198). Trotz dieser sehr einfachen Regel kann gezeigt werden, dass die mittlere Fehlerzahl im Zwei-Klassen-Fall höchstens doppelt so hoch wie die des Bayes-Klassifikators ist.

Eine weitere Verbesserung wird mit der Erweiterung der Entscheidungsregel erzielt: Die  $k$ -NN-Regel betrachtet  $k$  nächste Nachbarn und ordnet den unbekanntem Merkmalsvektor in diejenige Klasse ein, der die meisten der  $k$ -nächsten-Nachbarn angehören.

Zum zweiten Verfahrenstyp zählen die *linearen Klassifikatoren*. Vorgegeben wird eine Diskriminanzfunktion, im einfachsten Fall eine Hyperebene. Sind  $I$  Klassen vorgegeben, dann müssen  $I$  Entscheidungsebenen so plziert werden, dass sie eine optimale Trennung der Klassen erzielen. Als Maß für die Klassenzugehörigkeit wird eine gewichte Summe  $S_i(\mathbf{m})$  aus den Merkmalen für jeden Merkmalsvektor aufgestellt, Gl. (8-16). Wiederum ist die Entscheidungsfläche zwischen den Klassen  $K_i$  und  $K_j$  der Ort aller Punkte, für die  $S_i(\mathbf{m}) = S_j(\mathbf{m})$  gilt, so dass sich Hyperebenen mit der Normalengleichung Gl. (8-17) als Entscheidungsflächen ergeben. Die Gewichte der Diskriminanzfunktionen werden z.B. iterativ bestimmt, indem sukzessiv die vorliegenden Trainings-Instanzen klassifiziert werden. Jedes Mal, wenn ein hinzugenommener Merkmalsvektor falsch klassifiziert wurde, werden die Gewichte korrigiert.

$$S_i(\mathbf{m}) = w_{i,1} \cdot m_1 + w_{i,2} \cdot m_2 + \dots + w_{i,d} \cdot m_d + w_{i,d+1} \quad (8-16)$$

$$S_i(\mathbf{m}) - S_j(\mathbf{m}) = 0 \quad (8-17)$$

1. Fukunaga K: Introduction to Statistical Pattern Recognition; Academic Press 1990

Weichen die Verteilungen in den einzelnen Klassen sehr stark von Normalverteilungen ab (etwa bei Konkavitäten wie in Bild 8—6), dann reichen lineare Entscheidungsflächen möglicherweise nicht mehr aus. In solchen Fällen kann eine ungünstig verteilte Klasse in Unterklassen zerlegt werden. Da auch die Unter-Diskriminanzfunktionen linear sind, ergeben sich für die Trennung der Gesamtklassen *stückweise lineare* Entscheidungsflächen.

Komplexere Diskriminanzfunktionen lassen sich auf lineare Funktionen abbilden, indem der Merkmalsraum geeignet transformiert wird.

Das Prinzip der linearen Klassifikatoren liegt auch der *Support-Vektor-Maschine* (SVM)<sup>1</sup> zugrunde. Hier basiert die Konstruktion der Entscheidungsebenen lediglich auf wenigen Merkmalsvektoren, die den geringsten Abstand von der Ebene haben (die die Ebene „stützen“).

#### 8.4.3 Neuronale Netze

Neuronale Netze stellen einen speziellen Typ von überwacht lernenden Klassifikatoren mit nichtlinearen Diskriminanzfunktionen dar. Neuronale Netze sind aus miteinander vernetzten Verknüpfungselementen aufgebaut, die ein stark simplifiziertes Modell von Nervenzellen darstellen. Die Funktion des Grundelementes ist es, eine „weiche“ Entscheidung über eine gewichtete Summe von Eingangssignalen herzustellen, Bild 8—8. Üblicherweise wird als stetig differenzierbare Entscheidungsfunktion die *Sigmoid-Funktion*, Gl. (8–18) und Bild 8—7, gewählt.

Die Zusammenschaltung der Elemente folgt nun in der Regel einem Schichtenmodell: Alle Elemente einer Schicht sind mit jeweils allen Elementen der vorgeschalteten Schicht verbunden. Rückkopplungen sind zunächst nicht

1. Vapnik V: *Statistical Learning Theorie*, John Wiley & Sons, 1998;  
Vapnik V: *The Nature of Statistical Learning Theorie*, Springer, New York, 1995;  
Schölkopf B: *Support Vector Learning (Diss.)*, Oldenbourg Verlag, München, 1997

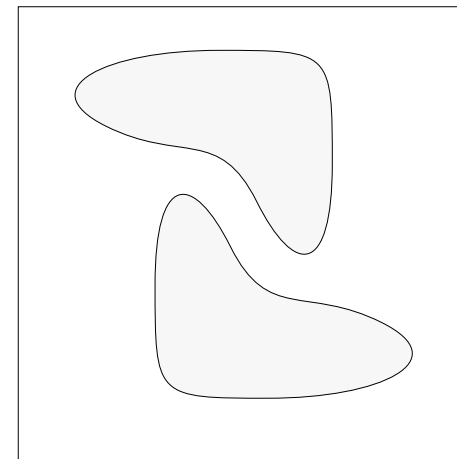


Bild 8—6

$$s(\mathbf{m}) = \frac{1}{1 + e^{-\frac{\mathbf{w} \cdot \mathbf{m} + a}{b}}} - \frac{1}{2} \quad (8-18)$$

$$\mathbf{w} \cdot \mathbf{m} = w_1 \cdot m_1 + w_2 \cdot m_2 + \dots + w_d \cdot m_d$$

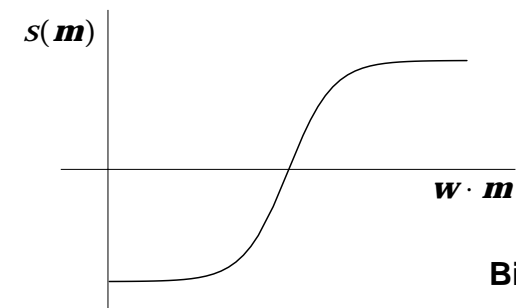


Bild 8—7

vorgesehen. Verbreitet sind drei-schichtige Netze mit einer Eingangsschicht aus den *Sensor-Neuronen*, einer versteckten Schicht ('hidden layer') aus den *Assoziations-Neuronen* und einer Ausgangsschicht aus den *Response-Neuronen*, Bild 8—9.<sup>1</sup> Die Sensor-Neuronen verteilen lediglich die Eingangssignale auf die Assoziations-Neuronen. Die Response-Neuronen stellen den eigentlichen Klassifikator dar: Der Eingangsvektor wird derjenigen Klasse zugeordnet, die durch das am stärksten ansprechende Response-Neuron repräsentiert wird.

Neuronale Netze stellen einen weitgehend heuristischen Klassifikator-Ansatz dar. Ihr Vorteil ist es, dass sie bei einheitlicher Grundstruktur infolge der Flexibilität ihrer Ausgestaltung (vor allem der freien Wahl der Anzahl von Assoziations-Neuronen) an sehr unterschiedliche Probleme angepasst werden können. Große Bedeutung kommt daher den Lernstrategien zu. Eine bewährte Strategie ist das Back-propagation-Training, ein Gradientensuchverfahren, bei dem das Fehlerquadrat zwischen tatsächlich erhaltenen und angestrebten Ausgangssignalen minimiert wird. Die Gewichte werden dabei während des Trainings von der Ausgangsschicht sukzessive in Richtung zur Eingangsschicht korrigiert, daher die Bezeichnung "backpropagation". Der bei Gradientenverfahren bestehenden Gefahr, in lokale Minima hinein zu laufen, kann durch Variation der Schrittweite ("Lernrate"), durch Erweiterung der Elementzahl in der versteckten Schicht und durch Wiederholung des Trainings mit unterschiedlichen Startgewichten begegnet werden.

Neuronale Netze (das Attribut "künstliche" davor zu stellen, wird meistens unterlassen...) haben in sehr kurzer Zeit eine hohe Attraktivität gewonnen, nachdem infolge der technologischen Fortschritte auch auf kleineren Rechenanlagen akzeptabel kurze Rechenzeiten erreicht werden konnten und mit applikations-spezifischen VLSI-Schaltungen sogar Hardware-Implementierungen neuronaler Netze realisiert worden sind. Ihr „Charme“ ist es, dass sie eine weitgehend generische Struktur besitzen, die durch Variation

1. Diese Struktur wurde bereits 1957 von *Rosenblatt* unter der Bezeichnung "Perceptron" vorgeschlagen.

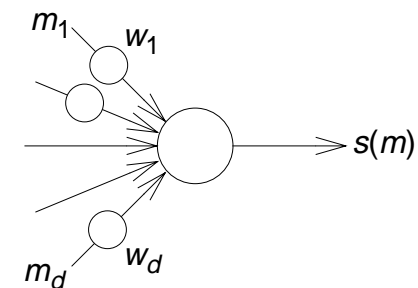


Bild 8—8

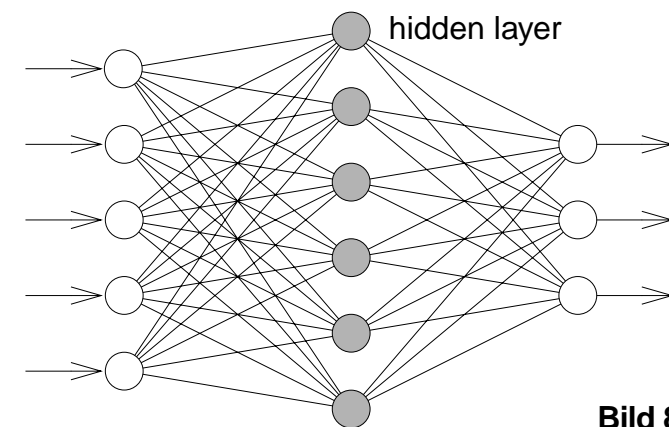


Bild 8—9

ihrer Parameter an sehr unterschiedliche Klassifikationsaufgaben angepasst werden kann. Ihr Nachteil ist, dass wir dabei keine Aufschlüsse über das Wesen der Aufgabe erwerben, denn das im Training des neuronalen Netzes erlernte “Wissen” bleibt implizit in den Gewichtsvektoren versteckt. Unsere Erfahrungen mit der Anwendung neuronaler Netze sind deshalb sehr schwierig in Regeln zu fassen. So erfolgt auch die Parametrisierung weitgehend nach der “trial-and-error”-Methode.

### 8.5 Unüberwacht lernende Klassifikatoren

Nicht bei jedem Problem können Trainingsmengen, d.h. Merkmalsvektoren klassifizierter Instanzen aller vorgegebenen Klassen in genügender Zahl gewonnen werden. Dies ist vor allem der Fall, wenn vor Auswertung des vorhandenen oder erwarteten Bildmaterials noch kein sinnvolles Klassensystem festgelegt werden kann. Typische Situationen liegen in der industriellen Qualitätssicherung beim Auffinden von Materialfehlern oder in der medizinischen Diagnostik vor, wenn über die Kausalität zwischen der Ursache und der bewirkten Bilderscheinung nur vage Hypothesen vorliegen.

Ein klassisches Verfahren eines unüberwacht lernenden Klassifikators ist die *Clusteranalyse*, die im nachfolgenden beschrieben werden soll.

Einen anderen interessanten Ansatz schlug *Kohonen* 1982 mit der sich *selbst organisierenden Merkmalskarte* (‘self-organizing feature map’), wiederum einem Netz aus künstlichen Neuronen, vor.<sup>1</sup> Durch ein inkrementelles Verändern der Gewichtungsfaktoren in den verbindenden “Axonen” derart, dass die Abstände zwischen den Gewichtsvektoren und den angebotenen Merkmalsvektoren bei jedem Iterationsschritt verringert werden, bildet das Neuronen-

---

1. Kohonen T: Self-organized Formation of Topologically Correct Feature Maps; Biological Cybernetics, Vol 43, 59 — 69



gitter eine Topologie aus, auf der sich die Verdichtungen einander ähnlicher Merkmalsvektoren im Merkmalsraum lokalisieren lassen.

### 8.5.1 Die Clusteranalyse

Ziel der Clusteranalyse ist es, in einer gegebenen Menge von Merkmalsvektoren eine Struktur zu finden, die eine Partition des Merkmalsraumes und damit eine Klasseneinteilung nahelegt. In diesem Wunsch ist impliziert, dass Merkmalsvektoren, die zu einer der sich so herausbildenden Klassen gehören, einander deutlich ähnlicher sind, als alle übrigen Merkmalsvektoren, dass die Merkmalsvektoren in dieser Weise also so genannte *Cluster*  $C_1 \dots C_j \dots C_I$ , mit den Schwerpunkten  $\hat{\mathbf{m}}_1 \dots \hat{\mathbf{m}}_j \dots \hat{\mathbf{m}}_I$  bilden. Merkmalsvektoren, die jeweils gemeinsam ein Cluster bilden, werden zu Instanzen einer Klasse  $K_j$  erklärt.

Als Voraussetzung müssen dazu zunächst ein Ähnlichkeitsmaß und ein Optimalitätskriterium gegeben sein. Bezüglich des Ähnlichkeitsmaßes ist es sinnvoll, alle Merkmale auf den Mittelwert Null und die Standardabweichung 1 zu normieren. Ein Optimierungsziel könnte die minimale Varianz der gefundenen Partitionen sein. Das Ähnlichkeitsmaß ist in diesem Fall der Abstand  $s$  nach Gl. (8–20) der Merkmalsvektoren vom jeweiligen Clusterschwerpunkt  $\hat{\mathbf{m}}_j$ , Gl. (8–19).

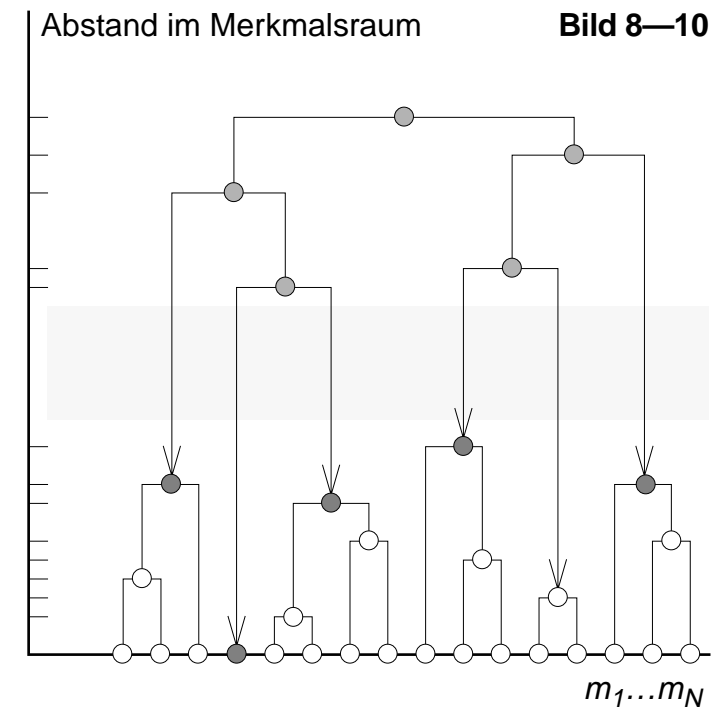
Häufig wird als Optimalitätskriterium die Summe  $s$  über den quadrierten euklidischen Abständen gewählt, Gl. (8–2). Dazu muss allerdings die Anzahl der Cluster bereits festliegen, da anderenfalls das Optimum darin bestünde, dass jedem Merkmalsvektor ein eigenes Cluster zugewiesen sein würde.

Wegen der hohen Zahl möglicher Partitionen ist ein direkter Weg, diejenigen zu finden, die das Optimalitätskriterium minimiert, praktisch nicht gegeben, insbesondere dann nicht, wenn a priori noch nicht einmal die Zahl der Cluster angegeben werden kann.

$$\hat{\mathbf{m}}_j = \frac{1}{N_j} \sum_{n=1}^{N_j} \mathbf{m}_{n,i} \quad (8-19)$$

$$s = \sum_{i=1}^I \sum_{n=1}^{N_i} |\mathbf{m}_{n,i} - \hat{\mathbf{m}}_i|^2 \quad (8-20)$$

Eine praktikable Lösung ist die *agglomerative Clusteranalyse*, ein hierarchisches Verfahren, das als Ergebnis das sogenannte *Dendogramm* liefert: Zu Beginn des Prozesses stellt jeder Merkmalsvektor ein eigenes Cluster dar. Die beiden ähnlichsten Cluster werden nun in einem neuen Cluster zusammengefasst, das auf der dem Abstand entsprechenden Ebene des Dendogramms eingetragen wird. Dieser Prozess wird rekursiv fortgesetzt, wobei der Schwerpunktsabstand der jeweils neu zusammengefassten Cluster wächst. Der Prozess ist abgeschlossen, wenn alle Merkmalsvektoren zu einem einzigen Cluster vereinigt worden sind. Nun kann das entstandene Dendogramm nach sinnvollen Partitionen analysiert werden, indem nach Abstandsniveaus gesucht wird, zwischen denen deutliche Abstandszunahmen auftreten, im dem in Bild 8—10 skizzierten Beispiel die schattierte Fläche. Die durch Pfeilspitzen gekennzeichneten Vereinigungen ergeben die gesuchten Cluster.



## 9 Bilddaten-Kompression

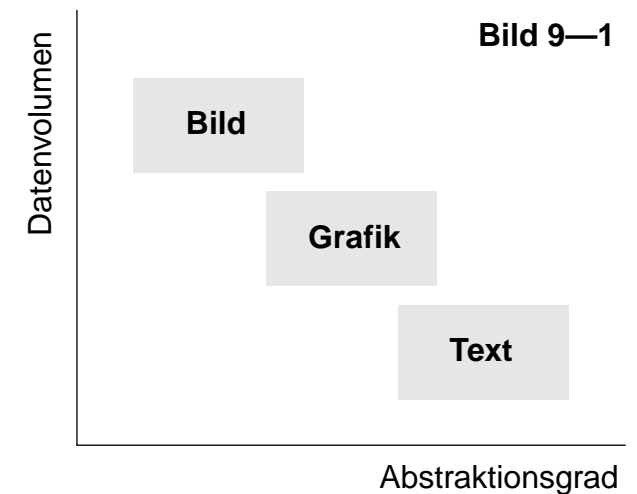
Beschreibungen einer realen oder imaginären Situation in Worten oder als Zeichnung erfordern eine wesentlich höhere Abstraktion als eine bildliche Darstellung (wenn eine solche möglich ist). Der Preis für den niedrigen Abstraktionsgrad eines Bildes ist dessen vergleichsweise enormes Datenvolumen, Bild 9—1:

Während eine Seite Text einige kByte oder eine Vektorgraphik einige zehn kByte für ihre Kodierung benötigen, liefert eine digitale Kamera mittlerer Qualität bereits Einzelbilder von 1 MByte Umfang.

Ca. 20 MByte werden es sogar, wenn die Ortsauflösung des Silberhalogenidfilms voll genutzt werden soll: 18 MByte für die Abtastung eines Kleinbildfilmes mit einer 3072·2048-Pixel-Matrix in drei Farbkanälen (PhotoDisc-Standard), 24 MByte für einen mit 4000·4000 Pixeln bei 12 bit Kontrastauflösung digitalisierten Röntgenfilm.

Dynamische Abbildungstechniken folgen noch weitgehend den Fernsehnormen (CCIR/PAL in Europa, NTSC in den USA): Die 575 aktiven Zeilen der CCIR-Norm ergeben bei einem Seitenverhältnis von 4:3 und 50 Halbbildern pro Sekunde einen Datenstrom von  $3 \cdot 575^2 \cdot 25 \cdot 4/3 \approx 33$  MByte pro Sekunde.

Seitdem Bilder digital verarbeitet und übertragen werden, ist die Suche nach kompakteren Kodierungen der Bildinformation ein aktuelles Forschungsthema. Wenn auch die Technologien für die Verarbeitung, Speicherung und Übertragung der umfangreichen Datenmengen bzw. Datenströme mittlerweile verfügbar sind, motivieren doch weiterhin die Kosten für Speicherkapazität und Übertragungsbandbreite. Ein starker Antrieb zur Standardisierung anwendungsspezifischer Verfahren lag zuletzt in der Minimierung lästiger Wartezeiten bei der Bildübertragung über die begrenzte Kanalkapazität des Internets.



Eine Betrachtung der Strategien und Verfahren zur Bildanalyse, insbesondere zur Segmentierung (Abschnitt 6 Seite 141), lässt darauf schließen, dass die von den Abbildungsprozessen gelieferten Primärsignale ein beträchtliches Maß an Redundanz aufweisen. Nun würde die vollständige Segmentierung eines Bildes erlauben, den Bilderhalt äußerst kompakt als Liste von Konturparametern und Merkmalen für die jeweiligen Segmentinhalte zu repräsentieren. Dies wurde in der Tat mit den sogenannten 'Bildkodierverfahren der zweiten Generation' erfolgreich realisiert, um beispielsweise Bewegtbilder reduzierter Qualität über 16kb/s-Funkkanäle zu übertragen. Derzeit sind weiterentwickelte Verfahren dieser Art Gegenstand von Standardisierungsbemühungen für Internet-Anwendungen.

Abgesehen von einer aufwendigen und nicht unproblematischen Verarbeitung, kann der Bildinhalt aus derart verdichteten Repräsentationen nicht wieder fehlerfrei rekonstruiert werden, denn eine Segmentierung setzt, wie gezeigt wurde, voraus, dass Bildstrukturen unterhalb einer vorzugebenden Stufe der Detailliertheit sowie die Ortsgenauigkeit ihrer Beschreibung für *irrelevant* erklärt werden dürfen. Der damit verbundene Informationsverlust kann zu rechtfertigen sein, wenn der Verwendungszweck oder das Auswertungsziel eine entsprechende Toleranz gegenüber Bildverfälschungen erlauben. Dies ist aber im allgemeinen damit verbunden, dass auch der Bildinhalt in engen Grenzen definiert ist.

Verfahren mit einer möglichst allgemeinen Anwendbarkeit und breiten Akzeptanz gehen daher von Signalmodellen anstatt von Modellen des Bildinhaltes aus. Wie bei der Segmentierung ist es auch in diesem Zusammenhang von Bedeutung, dass die abgebildeten Objekte und Szenen aus zusammenhängenden Regionen zusammengesetzt sein müssen, um einer gestaltorientierten Interpretation zugänglich zu sein. Weitere Fingerzeige gibt die Hypothese, dass der visuelle Wahrnehmungsprozess als Hierarchie aufeinander aufsetzender Verdichtungsprozesse gedeutet werden kann, deren primäre Stufe auf dem retinalen Reizmuster, dem Bildsignal aufsetzt. Wenn experimentell gezeigt werden konnte, dass das Bildsignal, das die

abgebildete natürliche Szene repräsentiert, Eigenschaften besitzt, die einen ersten Pixel-Verschmelzungsprozess unterstützen, kann auf ein beträchtliches Maß an Redundanz geschlossen werden. Herauszufinden ist also, worin die Redundanz einer Pixel-Matrix als Repräsentation einer (natürlichen) Szene besteht, um Transformationen entwerfen zu können, die zu weniger redundanten Repräsentationen führen.

Beschränken wir uns darauf, durch eine Signaltransformation lediglich die Redundanz vermindern zu wollen, dann kann jeweils eine *inverse* Transformation angegeben werden, mit der sich das primäre Signal *fehlerfrei* rekonstruieren lässt. Es überrascht nicht, dass Signaltransformationen, die redundanzvermindernd wirken, das Signal im allgemeinen auch in Komponenten zerlegen, die für die visuelle Wahrnehmung unterschiedliche Wichtigkeit (*Relevanz*) haben. Dies kann genutzt werden, um gegenüber einer ausschließlichen *Redundanzreduktion*, auch *verlustlose Kompression* (*lossless compression*) genannt, mit einer so genannten *Irrelevanzreduktion* oder *verlustbehafteten Kompression* (*lossy compression*) höhere Kompressionsgewinne zu erzielen.

Gegenüber verlustbehafteten Kompressionsverfahren, mit denen oft erheblich größere Kompressionsfaktoren erzielt werden als mit verlustfreien Kompressionsverfahren, ist jedoch Vorsicht angebracht, wenn eine sorgfältige, eventuell auch quantitative Auswertung erfolgen soll. Bei medizinischen Bildern sprechen sogar forensische Gründe gegen ihre Anwendung, da Rekonstruktions-Artefakte zu diagnostischen Fehldeutungen führen können.

### 9.1 Signal-basierte Kompression: Prädiktionskodierung

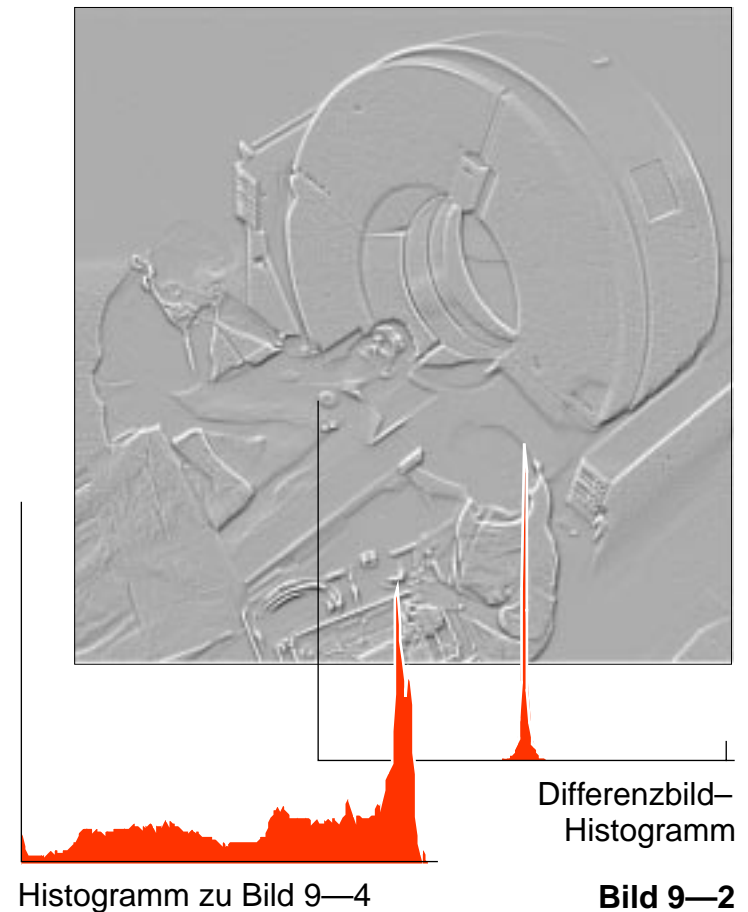
Wenn der Inhalt des zu komprimierenden Bildmaterials nicht bekannt ist, muss das gegebene Primärbild als stochastischer Prozess aufgefasst werden. Nun können im allgemeinen mindestens repräsentative Aussagen über die bildrelevanten Eigenschaften der abzubildenden Objekte oder Szenen sowie

über die abbildungsrelevanten Eigenschaften der Bildaufnahmekette gemacht werden. Wenn wir uns hier auf die Abbildung natürlicher Szenen und Objekte beschränken, gilt in einiger Allgemeinheit:

- Für ein beliebiges Pixel ist die Erwartung für eine homogene Nachbarschaft deutlich höher als für eine starke Fluktuation, denn Kanten und linienförmige Strukturen nehmen gegenüber homogenen Partitionen, die sie begrenzen, vergleichsweise geringe Flächenanteile ein. Daraus resultieren starke statistische Bindungen zwischen benachbarten Pixeln (siehe Verbundwahrscheinlichkeit, Gl. (6–6) auf Seite 147)
- Im Falle texturierter Partitionen sind entsprechende statistische Bindungen auf der Ebene texturbeschreibender Merkmale zu erwarten.
- Abbildungsprozesse haben physikalisch bedingt grundsätzlich einen Tiefpasscharakter, der Sprünge der primären Signalgröße glättet und damit die Signaldifferenz zwischen benachbarten Pixeln reduziert.

Deutlich kommen diese Signaleigenschaften in Histogrammen über differenzierte Bilder zum Ausdruck, die im allgemeinen eine starke Konzentration bei Null aufweisen. Bild 9–2 zeigt das Differenzbild von Bild 9–4 (Seite 216 oben) und die Histogramme beider Bilder. Aus den Differenzbild-Histogrammen repräsentativer Bilder lässt sich auf eine hohe Auftrittswahrscheinlichkeit kleiner Differenzwerte und ein entsprechend seltenes Auftreten größerer Werte schließen. Informationstheoretisch gesprochen sinkt die Entropie mit der wachsenden Amplitude der Differenzwerte; entsprechend steigt der Informationsgehalt. Dies trifft sich mit unserer Hypothese über die Wichtigkeit von Kanten- und Linienstrukturen für die Interpretation des Bildinhaltes.

Die Informationstheorie gibt auch eine Antwort auf die Frage einer optimalen Kodierung: Die benötigte Anzahl der Bits für ein Informationsinkrement (hier ein Pixel des Differenzbildes) ist durch den Informationsgehalt des Ereignisses, d.h. durch den negativen Logarithmus seiner Auftrittswahrscheinlichkeit gegeben, Gl. (9–1). Der mittlere Informationsgehalt, die



$$I(g_j) = -\log_2[p(g_j)] \quad (9-1)$$

$$H(g) = -\sum_j p(g_j) \cdot \log_2[p(g_j)] \quad (9-2)$$

Entropie nach Gl. (9-2), würde minimiert werden, wenn es gelänge, den Differenzamplituden Kodewörter unterschiedlicher Länge zuzuordnen, deren Bitlängen den Informationsgehalt möglich gut approximieren. Bei der Konstruktion solcher *Optimalkodes* muss berücksichtigt werden, dass die einzelnen Kodewörter trotz ihrer variablen Bitzahl eindeutig voneinander separiert werden können. Kodfamilien, die diese Eigenschaft besitzen, wurden von *Shannon* und *Fano* sowie von *Huffman* bereits 1952 entwickelt.<sup>1</sup>

Kompressionsverfahren dieser Klasse werden je nach Sichtweise als *DPCM* ('*differential pulse code modulation*') oder als *Prädiktionskodierer*<sup>2</sup> bezeichnet. Ihre Realisierung ist wenig aufwendig: Die senderseitige Transformation erfolgt durch die Faltung mit einem *kausalen* Differenzoperator, gefolgt von einem Optimalkodierer. Auf der Empfängerseite wird das dekodierte Differenzsignal durch ein zur Differenzierung inverses Filter in den Bildraum zurücktransformiert, Bild 9-3.

Im einfachsten Fall besteht die Transformation in einer Differenzbildung zwischen zwei benachbarten Pixeln. Die Rücktransformation ist in diesem Fall eine Summation des bereits rekonstruierten Nachbar-Pixels mit dem dekodierten Differenzwert. Effektiver sind Differenzoperatoren, die die Nachbarschaft in beiden Dimensionen berücksichtigen, z.B. Gl. (9-3).

Problematisch am Prädiktionskodierer ist die verfahrensbedingte Akkumulation von Übertragungsfehlern. Nachteilig ist überdies ihr geringer Kompressionsfaktor bei stark texturierten Bildinhalten: Redundanz, die durch die Gleichförmigkeit sich wiederholender Detailstrukturen entsteht, kann auf diesem auf einfachen Weg nicht separiert werden.

1. Der Begriff 'optimal' wird hier im Sinne einer *Quellenkodierung* benutzt. Das Kriterium ist die Minimierung des Datenvolumens für die Darstellung einer Information.
2. Der Empfänger schätzt den zu rekonstruierenden Bildpunkt auf der Basis bereits rekonstruierter Nachbarpunkte und korrigiert den Prädiktionswert durch den aktuell übertragenen Differenzwert, die 'Innovation'.

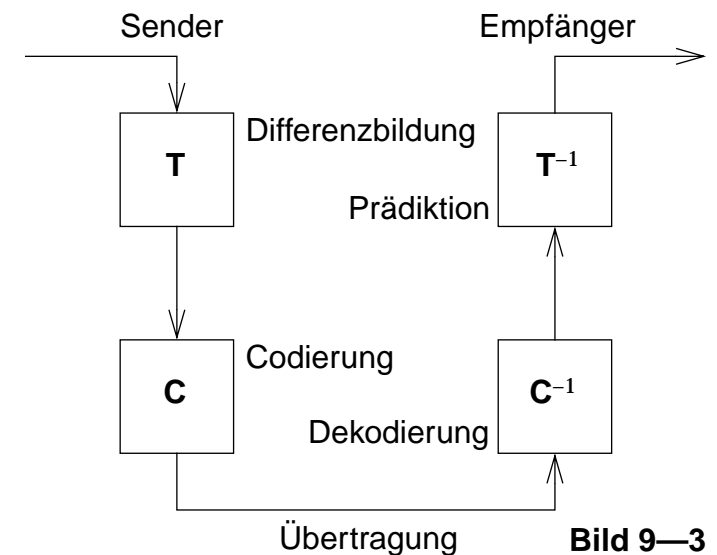


Bild 9-3

$$h(m, n) = \frac{1}{27} \begin{bmatrix} -10 & +27 \\ -7 & -10 \end{bmatrix} \quad (9-3)$$

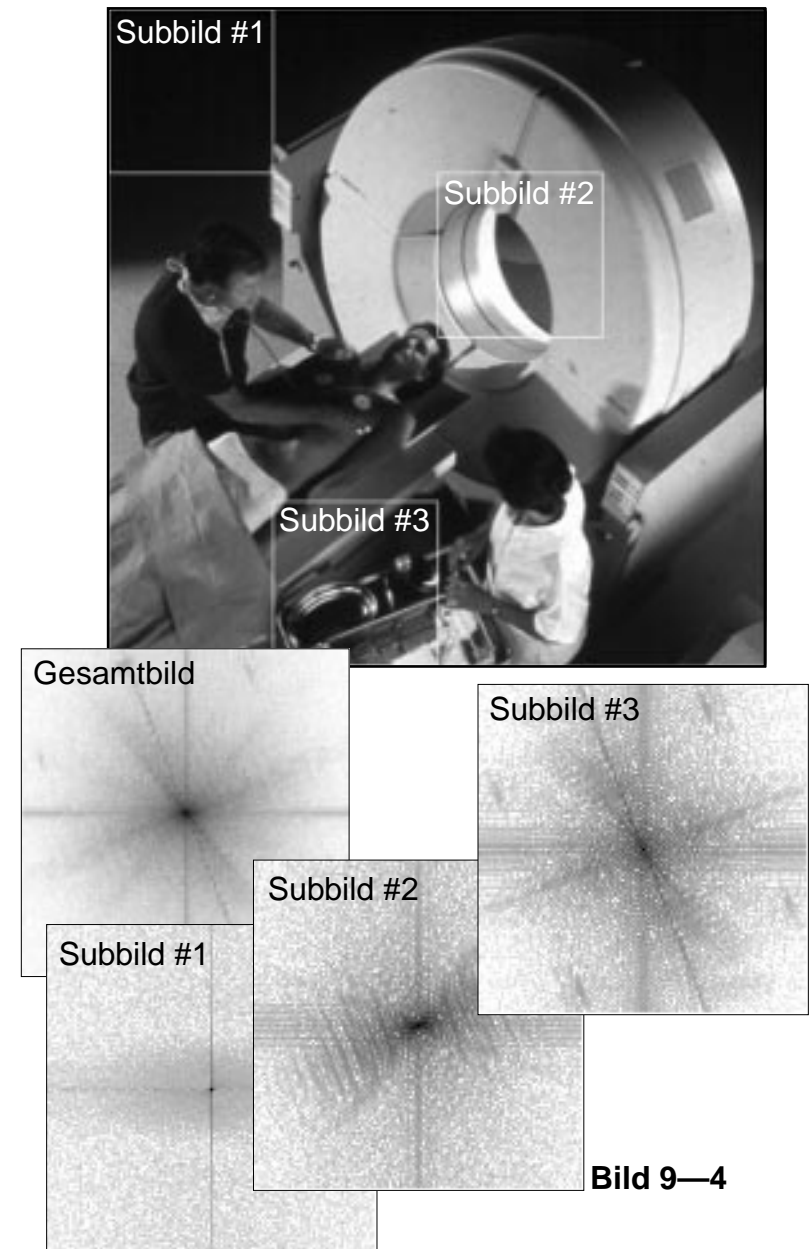
## 9.2 Spektrum-basierte Kompression: Transformations-Kodierung

Greifen wir deshalb noch einmal die Frage auf, worin sich die Redundanz in konventionellen Bildern verkörpert. Einen komplementären Einblick liefert eine spektrale Betrachtung: Zu beobachten ist in Leistungsdichtespektren grundsätzlich eine starke spektrale Leistungskonzentration um den Ursprung herum: Großflächige, weitausgreifende Fluktuationen führen zu starken Beiträgen niedriger Wellenzahlen. Konturen- und Linienstrukturen liefern Beiträge bei hohen Wellenzahlen, deren Amplituden jedoch wegen ihrer geringen lokalen Ausdehnung klein sind. Texturen liefern ihrem Wesen entsprechend Beiträge, die im Spektralbereich enger lokalisiert sind.

Die spektrale Betrachtung macht besonders deutlich, dass der primär angebotene Übertragungskanal, d.h. Bandbreite mal Anzahl der Quantisierungsstufen, sehr schlecht ausgenutzt wird, wenn jedem Spektralkoeffizienten der maximale Amplitudenbereich eingeräumt wird. Nun dürfen auch Koeffizienten mit sehr kleiner Amplitude nicht einfach unterdrückt werden, denn diese verkörpern ja gerade die für die Interpretation wichtige Strukturinformation. Naheliegend ist deshalb die Anwendung einer Optimalkodierung im Spektralbereich. Der Empfänger hat eine entsprechende Dekodierung mit anschließender Rücktransformation vorzunehmen.

Deutlich steigern lässt sich die Effizienz einer *Transformations-Kodierung*, wenn das zu komprimierende Bild soweit in Subbilder unterteilt wird, dass die einzelnen Subbilder vorwiegend in sich homogene oder homogen texturierte Teilbereiche erfassen, Bild 9—4. Dann kann das Kodierschema sogar an den spektralen Aufbau der jeweils individuell transformierten Subbilder adaptiert werden (häufig gewählte Subbildgrößen sind 8·8 oder 16·16 Pixel).

Anstelle der diskreten Fourier-Transformation hat sich die *diskrete Cosinus-Transformation* (DCT) durchgesetzt (siehe Abschnitt 5.5.3 Seite 127). Zum einen befreit die DCT von komplexer Zahlenrechnung. Zum anderen verursachen Rekonstruktionsfehler infolge gestörter oder ungenau kodierter Koeffizienten subjektiv geringer störende Artefakte, da Cosinus-Funktionen





auf Grund ihrer Symmetrie kontinuierliche Übergänge an den Subbildrändern ergeben (bis auf die Grundwelle, da diese bei der Rekonstruktion wegen der Spiegelung der Signalfunktion wie alle Wellen zu halbieren ist). Ein vollständiges System besteht aus Transformationstufe, Kodierer, Dekodierer und Rücktransformationstufe (inverse DCT), Bild 9—5.

### 9.3 JPEG — Beispiel eines verbreiteten Standards

Unzählige Kompressionsverfahren, optimiert nach unterschiedlichen Kriterien wie Kompressionsrate, Rechenaufwand, Fehlerempfindlichkeit oder spezialisiert auf bestimmte Anwendungen oder Bildinhalte, wurden publiziert, seit digitale Bildübertragung und -speicherung für professionelle Anwendungen interessant wurden. Verbreitung haben allerdings nur die wenigen Verfahren gefunden, die es zu standardisieren gelang. Hier spielte das Internet eine entscheidende Rolle. Bildübertragung ohne Datenkompression würde zu völlig unakzeptablen Übertragungszeiten führen. Ohne Standardisierung wiederum ist ein offenes Kommunizieren nicht denkbar.

Ein Standard, der sich vor allem im Internet durchgesetzt hat, ist das von der *Joint Photographics Experts Group* (JPEG) eingebrachte gleichnamige Kompressionsverfahren.<sup>1</sup> Bei diesem Standard werden alternativ zwei unterschiedliche Wege gewählt, je nachdem, ob eine verlustlose Kompression gefordert wird oder ob eine höhere Kompressionsrate unter Inkaufnahme von Rekonstruktionsfehlern gewünscht wird. Im letzten Fall kann das Optimierungskriterium überdies zwischen der Minimierung einer subjektiven Bildqualitätseinbuße und der Maximierung der Kompressionsrate verschoben werden. Im folgenden beschränken wir uns auf die Anwendung des JPEG-Standards auf Grauwertbilder.

1. Pennebaker WB, Mitchell JL: JPEG — Still Image Compression Standard; Van Nostrand, New York 1993

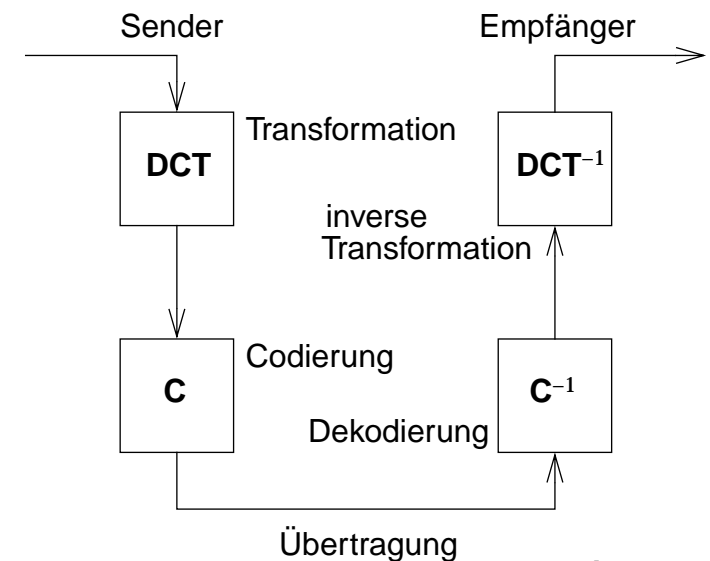


Bild 9—5

## 9.3.1 'Lossles JPEG'

Für die verlustlose Kompression wird ein DPCM-Verfahren mit einem Entropie-Koder eingesetzt. Für die Kodierung stehen unterschiedliche 'Kodebücher' zur Auswahl. Im Kopfeintrag des komprimierten Bildes findet der Empfänger alle Angaben, die für die Dekodierung benötigt werden.

## 9.3.2 'Lossy JPEG'

Für den Fall, dass Rekonstruktionsfehler zulässig sind, wird eine Kombination aus DPCM und DCT angewandt. Die DCT wird mit einer Blockgröße von 8·8 Pixeln durchgeführt, liefert also eine 8·8-Matrix  $G(0..7, 0..7)$ . Der Koeffizient  $G(0,0)$ , der mittlere Grauwert des Subbildes, muss sehr präzise rekonstruiert werden, damit die Schachbrettstruktur der Subbildaufteilung nicht sichtbar wird. Da große Grundhelligkeitsdifferenzen auch über die Subbilddistanz hinweg weniger wahrscheinlich sind, werden die  $G(0,0)$ -Koeffizienten benachbarter Subbilder nach einem DPCM-Schema kodiert.

Die übrigen 63 Koeffizienten eines Subbildes werden in 64 Amplitudenstufen quantisiert und nach einem diagonal ausgerichteten Zickzack-Schema zu einer eindimensionalen Sequenz umgeordnet, innerhalb der die Koeffizienten näherungsweise nach ihrer radialen Wellenzahl geordnet sind, Bild 9—6. Die quantisierten Koeffizienten werden anschliessend mit einer Kombination aus Lauflängen-Kodierung und Entropie-Kodierung weiter komprimiert. Sowohl die Quantisierungskennlinie als auch das Kodebuch werden als Tabellen festgelegt und dem Datensatz des kodierten Bildes vorangestellt. Auf diese Weise ist eine flexible Anpassung der Kompressionseigenschaften (Bildtyp, Rekonstruktionqualität) an die Anwendung möglich.

Noch bessere Ergebnisse hinsichtlich des Kompressionsgewinnes und der Rekonstruktionsartefakte liefert die Wavelet-Transformation (WT), siehe folgende Seite. Die WT liegt der jüngsten Variante *JPEG 200* zu Grunde.

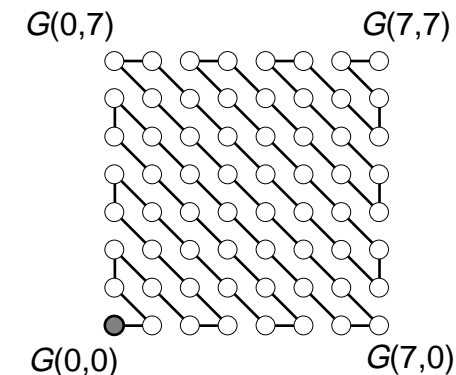


Bild 9—6

#### 9.4 'Subband coding' — Wavelet-Transformation

Das JPEG-Verfahren bietet einen interessanten Nebeneffekt: Mit einem sehr kompakten Datensatz lediglich aus den  $G(0,0)$ -Koeffizienten kann bereits ein grobes Übersichtsbild aufgebaut werden. Genutzt werden kann dies, um einen schnellen Überblick über Bildinhalte in einer Bilddatenbank zu erhalten, ohne jeweils den kompletten Datensatz auslesen zu müssen. Angenehm ist ein frühzeitiger grober Bildeindruck auch bei einer langsamen Bildübertragung. Der progressive Aufbau des Bildes beim Empfänger (*progressive reconstruction*) hat in diesem Fall gegenüber einem zeilenweisem Aufbau (*sequential reconstruction*) erhebliche psycho-visuelle Vorteile.

Für einen optimal progressiven Bildaufbau ist die Zerlegung des Bildes in eine *Pyramide* abgestufter örtlicher Auflösung wünschenswert. Perfektioniert wird die dazu erforderliche spektrale Zerlegung des Bildinhaltes und der stufenweise Wiederaufbau in der Reihenfolge der Spektralkomponenten mit Kodierungstechniken, die die *Wavelet-Transformation* (WT) als Basis verwenden.

Die spektrale Dekomposition der Ortsfunktion des Bildes wird mit einem Satz separabler Filter durchgeführt, der eine Familie reeller orthogonaler Basisfunktionen bildet, die 'wavelets'. Die besondere Eigenschaft dieser Filter ist es, dass sie durch Skalierung und Translation aus einem sogenannten 'mother wavelet' entwickelt werden können.

Als 'mother wavelet' wird ein Filter konstruiert, das den Spektralbereich „weich“ halbiert, wie es das Beispiel in Bild 9—7 veranschaulicht. Dabei muss ein Kompromiss zwischen spektraler Trennschärfe und Lokalität (örtliche Ausdehnung) des Filterkernes gemacht werden. Aus diesem Tiefpassfilter wird ein komplementäres Hochpassfilter entwickelt (gepunktete Kurve). Mit dem Filterpaar wird das Signal in zwei sich ergänzende Spektralbänder zerlegt — zweidimensionale Signale entsprechend in vier Spektralbereiche.

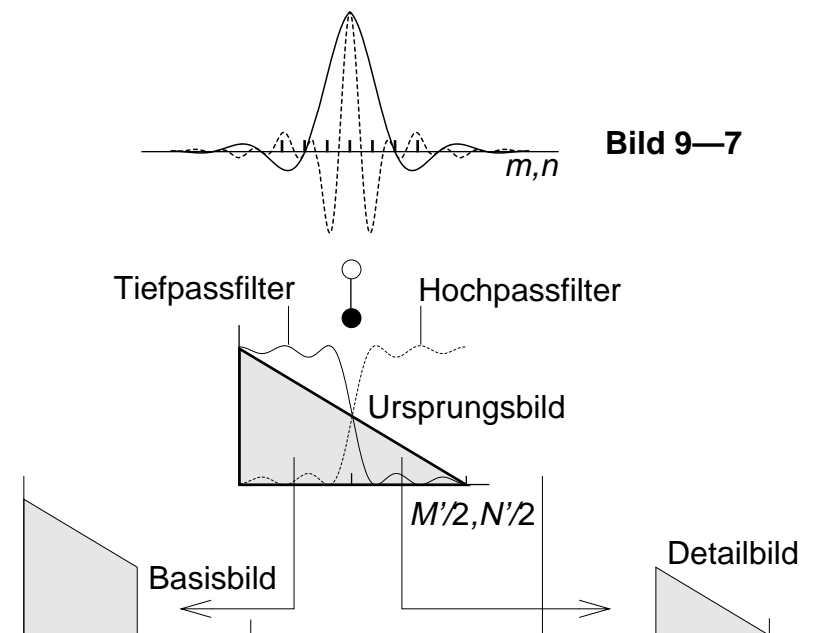
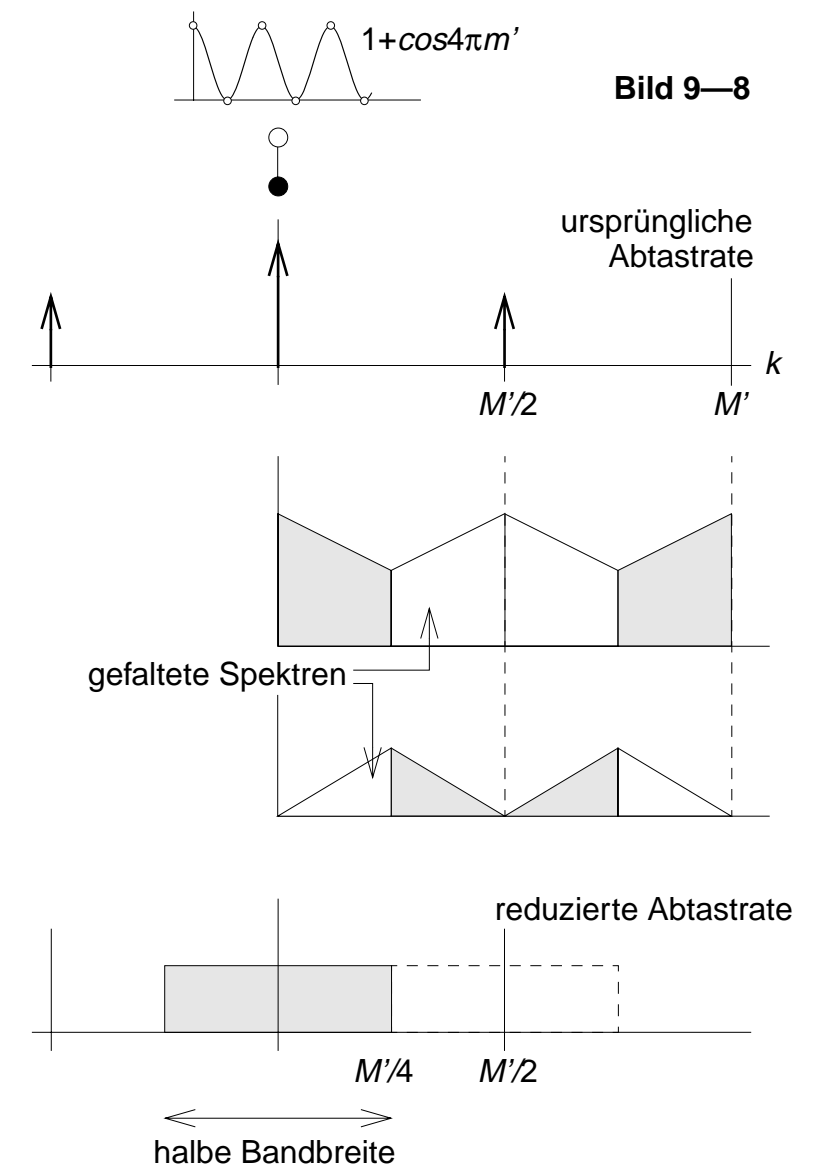


Bild 9—7

Die Abtastrate kann aufgrund der Halbierung der Bandbreite der gefilterten Signale in beiden Dimensionen halbiert werden, so dass vier Teilbilder von jeweils einem Viertel der ursprünglichen Pixelzahl erzeugt werden. Die Halbierung der Abtastrate kann als Multiplikation mit einer Welle  $k = M'/2$  beschrieben werden, die jedes zweite Pixel löscht ('decimation'), Bild 9—8. Im Spektralbereich führt die Faltung beider gefilterten Signale mit der Transformierten der Cosinus-Welle zu einer Verdopplung der Wiederholrate der Signalspektren, d.h. zu einer Halbierung der Bandbreite. Die so erzeugten vier Teilbilder repräsentieren ein vergrößertes Eingangsbild sowie die daraus herausgelösten horizontalen, vertikalen und diagonalen Detailstrukturen. Die spektrale Zerlegung fügt also keine Redundanz hinzu.

Die WT zerlegt das Bild stufenweise, Bild 9—9: Rekursiv wird die spektrale Zerlegung durch horizontale bzw. vertikale Tiefpass/Hochpass-Filterpaare auf das jeweilige Basisbild angewandt. Die anschließende Halbierung der Abtastrate ( $\downarrow_2$ ) entspricht dabei einer Skalierung der Filter der nächsten Stufe mit dem Faktor  $1/2$ . So ergibt sich eine Pyramide von Subbildern, in die der Spektralgehalt des Bildes zerlegt worden ist (Bild 6—11 auf Seite 153 zeigt ein Beispiel einer solchen Zerlegung). Der Kodierprozess kann nun für jedes Subbild individuell optimiert werden.

Die Anwendung der WT für die Bilddatenkompression setzt die Existenz einer *inversen WT* voraus, denn hier soll das Bildsignal nicht nur wie in der auf Seite 153 genannten Anwendung für die Texturanalyse (Abschnitt 6.1.5) spektral zerlegt werden. Bis auf Restfehler, die von der Konstruktion des 'mother wavelet' abhängen, kann das Signal aus den Teilspektren wieder zusammengesetzt werden. Skalierung und spektrale Translation müssen dazu wieder rückgängig gemacht werden: Beginnend auf der untersten Stufe, dem Basisbild, werden die Abtastraten der Teilbilder verdoppelt ( $\uparrow_2$ ) und diese über Tiefpass/Hochpass-Filterpaare zusammengefügt, Bild 9—10. Den Filtern kommt die Rolle zu, die zur Verdopplung der Abtastfrequenz einzufüllenden Bildpunkte zu interpolieren, indem die bei der Verdopplung der Abtastfrequenz auftretende Spektrumsspiegelung unterdrückt wird.



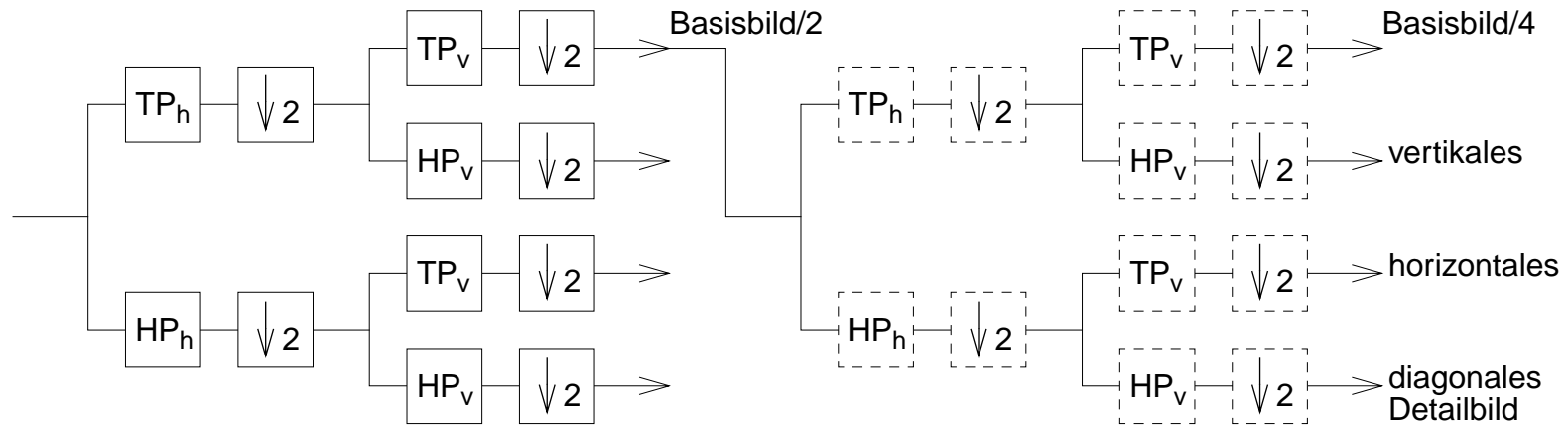


Bild 9—9

Dieser Prozess wird wiederum rekursiv fortgesetzt, bis die ursprüngliche Bildmatrix erreicht ist. Auf jeder Stufe der Rekonstruktion liegt bereits ein Bild der jeweiligen Auflösungsstufe vor, das insoweit korrekt ist, als Artefakte infolge der jeweiligen Unterabtastung durch den durchlaufenen Filterprozess mit approximierten Ideal-Tiefpässen und -Hochpässen minimiert worden sind.

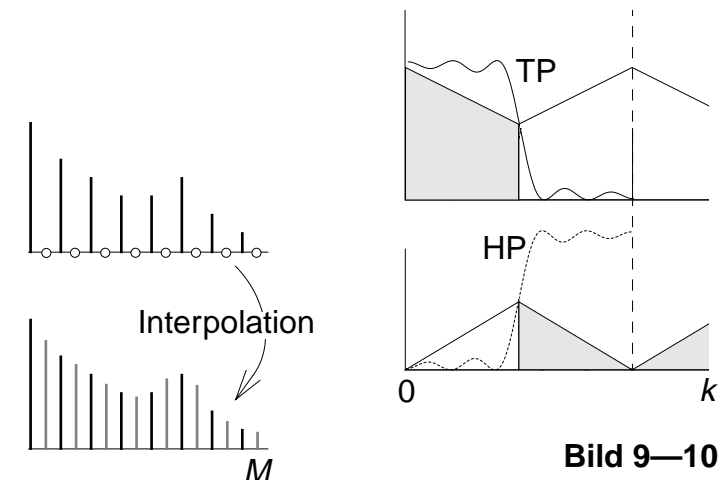
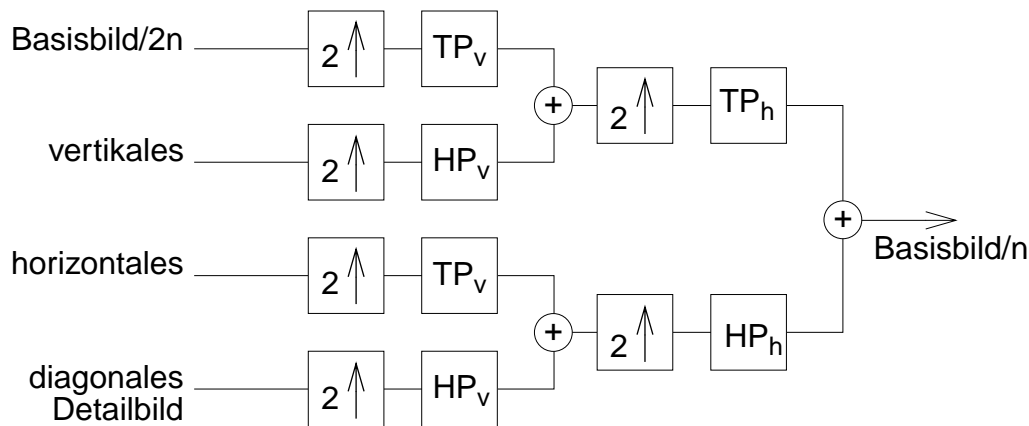


Bild 9—10



# Stichworte

## A

Abbildung 55  
Abbildungen  
  –modell, geometrisches 16  
  –prozess 11  
Absorptionskoeffizient 82  
Abstand 198, 201  
Abtast  
  –rate 52  
  –theorem 52  
Ähnlichkeitssatz 51  
akustische Linse 74  
Artefakt 49  
Autokorrelationsfunktion (AKF) 124

## B

backprojection 85  
Bayes 200  
Bild  
  Projektions– 81  
  –raum 13  
  –sensor 55  
  –signalfunktion 38  
Binomial  
  –Funktion 115  
  –koeffizienten, Tabelle 115  
  –verteilung 27  
Bloch–Gleichung 93

## C

Camera, siehe Kamera  
Canny–Detektor 162  
CCD  
  frame–transfer 67  
  full–frame 66  
  interline 67  
  –Matrixsensor 63  
  –Schieberegister 64  
Clusteranalyse 209  
Computer–Tomographie 82  
Computer–Visualistik 3  
cooccurrence matrix 148  
Cosinus–Transformation (DCT) 127, 216

## D

Delta  
  –funktion, 2D 47  
  –gerade 47  
Dendogramm 210  
Dephasierung 96  
Dilatation 191  
Diskretisierungsnetzwerk 176  
Diskriminanzanalyse 204  
Dispersion 23  
Distanz 198  
  –Funktion 176, 193  
DPCM 215

## E

Eigenfunktion 38  
Enhancement 101  
Entropie 215  
  Verbund– 148  
Entscheidungsregel 200  
Erosion 191  
Erwartung 23  
Eulerscher Satz 38

## F

Faltung 32, 36, 51  
fanbeam 85  
feature 143  
Filter  
  Bandpass– 116  
  FIR– 110  
  Hochpass– 116  
  IIR– 117  
  Kanten– 151  
  Median– 120  
  Optimal– 121  
  rekursives 117  
  Textur– 151  
  Wiener– 121  
Finite...  
  Differenzen 183  
  Elemente (FEM) 184

Fokussierung  
  dynamische 74  
Fourier  
  –Deskriptor 181  
  –Transformation 36  
  –Transformation, 2D– 42

Freeman code 182

Funktion  
  2D–Punkt– 47  
  Autokorrelations– 124  
  Binomial– 115  
  Distanz– 176  
  Eigen– 38  
  Gauß– 23  
  scha– 47  
  Shape–density– 180  
  si– 45  
  Wellen– 38

Projektionsfunktion 53

## **G**

Gammakamera 71, 99  
Gauß–Funktion 23, 46  
  2D– 46  
Geometrie  
  diskrete 176  
  euklidische 16  
Gestalt 174  
  –gesetz 157  
  –kriterien 187  
Gradient 149  
Graph 176

Grenzkontrast 24, 25  
gyromagnetische Konstante 93

## **H**

Halbbildverfahren 63  
Hamming–Fenster 153  
Harmonisierung 112  
Hauptachsen–Transformation 204  
Histogramm 107, 214  
  –Equalisierung 106  
  –Equalisierung, adaptive 108  
  kumulatives 107  
Homogenitätsmaß 168  
Hough–Transformation 188  
Hounsfield–Skala 88  
Huffman–Code 215

## **I**

Impuls  
  2D–Delta– 31  
  –antwort 29  
Instanz 196  
interlace scan 63  
Interpolation  
  bi–kubische 131  
  bi–lineare 130  
  Spline– 131  
Irrelevanzreduktion 213

## **J**

JPEG 217

## **K**

Kamera  
  Fernseh– 59  
  Gamma– 71, 99  
  Loch– 19  
Kante 158  
Kanten  
  –Detektor 159  
  –filter 151  
  –Relaxation 164  
Karhunen–Loève–Transformation 204  
Kernspin–Tomographie 92  
Klassifikation 195  
Kodierung  
  Prädiktions– 213  
  Transformations– 216  
Kollimator 71  
Kompression 211  
  verlustbehafteten 213  
  verlustlose 213  
Konnektivität 176  
Kontrast 24, 148  
Kontrasttransformation 105  
Kontur 159, 163  
  aktive 190  
  diskrete 182  
  kontinuierliche 179  
  –Suche 187  
Koordinaten 16  
  homogene 16  
Kovarianz 148  
Kreuzkorrelation 151



## L

Lamor-Frequenz 94  
Laplace-Operator 150  
Laserscanner 68  
Leistungsdichte-Spektrum 124, 152  
Lernen  
    überwachtes 197, 203  
    unüberwachtes 208  
linearphasig 49  
Linie  
    Linienpaar, lp/mm 39  
LSI-System 29

## M

Magnetresonanz-Tomographie (MR) 92  
Mahalanobis-Abstand 201  
Manhattan-Abstand 198  
Matrix  
    Grauwertübergangs- 148  
    Transformations- 17  
Medianfilter 120  
Merkmal 143  
    spektrales 152  
    statistisches 146  
    Struktur- 149  
Merkmals...  
    -raum 197  
    -vektor 154, 197  
Minkowski-Addition, -Subtraktion 192  
Mittelwert 147  
modulation transfer function (MTF) 48

## morphologische

    Öffnung 192  
    Operationen 191  
    Schließung 192  
Muster 197

## N

Neuronales Netz 206  
Nipkow-Scheibe 59  
Normalverteilung 23  
Nuklearmedizin 71, 99  
nullphasig 49, 118

## O

Operation  
    dyadische 102  
    geometrische 128  
    globale 125  
    lokale 110  
    monadische 103  
    morphologische 191  
    Punkt- 105  
Operator  
    Gradienten- 149  
    Laplace- 150  
    Sobel- 151  
optical transfer function (OTF) 48  
Optimal...  
    -filter 121  
    -kodierung 215

## P

PET 100  
phase transfer function (PTF) 48  
phased array 74  
Phasen  
    -kodierung 97  
    -winkel 39  
photo multiplier tube (PMT) 72  
photoelektrischer Effekt 59  
Photogate 64  
Pixel 21, 30  
point spread function (PSF) 30  
Prädiktionskodierung 213  
Präzession 94  
principal component analysis 204  
progressive reconstruction 219  
Projektion 95  
    1D- 82  
    Parallel- 82  
    perspektivische 19  
    Zentral- 19, 85  
Projektionsfunktion 53  
Protonenspin 93  
Prototyp 196  
Punkt  
    -funktion 47  
    -raster 47

## Q

quad tree 170

Quanten

–fluenz 28

–fluss 28

–rauschen 27

## R

Raster

hexagonales 177

pseudo–hexagonales 178

Rauschen 22

Photonen– 27

Quanten– 26

Quantisierungs– 26

Speckle– 75

thermisches 27

rect–Funktion 45

Redundanzreduktion 213

region growing 168

Rekonstruktion, 3D– 77

Relaxation

Kanten– 164

Kernspin– 94

Restauration 101

Röntgenstrahlung 27, 81

root–mean–square, rms 23

Rotation 18, 51

Rückprojektion 87

Runge–Gleichung 40

## S

Scanning–Laser–Ophthalmoskop 69

Schätzung 23

Schwellwert 166

Segment 142

Segmentierung 142

Kanten–orientierte 155

Regionen–orientierte 166

Separierbarkeit 43

si–Funktion 45

signal-to-noise ratio (SNR) 29

Skalierung 18

Skelettierung 193

Sobel–Operator 151, 159

Speckle 75, 144

SPECT 99

Spektrum 40

Leistungsdichte– 124, 152

Spiegelung 51

Spline 131, 184

split and merge 169

Standardabweichung 23, 147

Statistik 147

subband coding 219

Symmetrie 51

Systemtheorie 35

Szintillator 71

## T

Filter

Textur– 151

Textur 144

Theorem

der Fourier–Transformation 50

Abtast– 52

Zentralschnitt– 53

Tomographie

Computer– (CT) 82

MR– 92

Transducer

phased–array– 74

Ultraschall– 74

Transformation

Cosinus– 127, 216

Fourier– 36

geometrische 17

Hough– 188

Wavelet– 153, 219

Transformations–Kodierung 216

Translation 18

## U

Übertragungsfunktion 40, 48

Ultraschall–Scanner 72

Unschärfe 29

–Relation 153

unsharp masking 114

## **V**

Varianz 23, 147  
Vektor 16  
    Merkmals- 154  
Verschiebungssatz 51  
Verzerrung  
    geometrische 129  
    perspektivische 18  
Vidicon 59  
Visualisierung 8  
Visualistik 1  
Voxel 82, 88

## **W**

Wahrscheinlichkeit 200  
Wasserscheiden-Algorithmus 165  
Wavelet-Transformation 153, 219  
Welle...  
    harmonische 44  
    komplexe 38, 92  
Wellen...  
    -funktion 38  
    -zahl 39  
Wiener-Filter 121

## **Z**

Zeilensprungverfahren 63  
Zentralprojektion 19  
Zentralschnitt-Theorem 53, 83

