



Handbuch der Informatik

herausgegeben von

Prof. Dr. Albert Endres

Prof. Dr. Hermann Krallmann

Dr. Peter Schnupp

Band 6.4

Verarbeiten und Verstehen von Bildern

herausgegeben von
Prof. Bernd Radig
TU München

mit Beiträgen von
Prof. Dr. Paul Levi
Wolfgang Eckstein
Uwe Meyer-Gruhl
Josef Pauli
Karlhorst Klotz
TU München

R. Oldenbourg Verlag München Wien 1993

Prof. Dr. Bernd Radig

Studium der Physik an der Universität Bonn, 1978 Promotion, 1982 Habilitation, beides an der Universität Hamburg; im Anschluß Professur an der Universität Hamburg. Seit 1986 Professor für Informatik an der Technischen Universität München. Arbeitsgebiete: Kognitive und Wissensbasierte Systeme, Künstliche Intelligenz und Bildverarbeitung. 1988 Gründungsdirektor und Sprecher des Bayerischen Forschungszentrums für Wissensbasierte Systeme (FORWISS).

Anschrift:
TU München
Institut für Informatik
Orleansstr. 34
81667 München

email:radig@informatik.tu-muenchen.de

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Handbuch der Informatik : [die umfassende Darstellung der Informatik in Einzelbänden] / hrsg. von Albert Endres ... - München ; Wien : Oldenbourg.

NE: Endres, Albert [Hrsg.]

Bd. 6.4. Verarbeiten und Verstehen von Bildern / hrsg. von

Bernd Radig ... - 1993

ISBN 3-486-20787-3

NE: Radig, Bernd [Hrsg.]

© 1993 R. Oldenbourg Verlag GmbH, München

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Gesamtherstellung: Huber KG, Dießen

ISBN 3-486-20787-3

Inhaltsverzeichnis

Vorwort der Reihen-Herausgeber	9
Danksagung	11
1. Einleitung	13
2. Grundlagen und Begriffsdefinitionen des Bild- verstehens	21
2.1 Schichtenmodell des Bildverstehens.....	22
2.1.1 Deklarative Wissens Ebenen des Bildverstehens.....	22
2.1.2 Prozedurale Wissens Ebenen des Bildverstehens.....	33
2.2 Kontinuierliche und digitale Bilder	35
2.3 Lineare und verschiebungsinvariante Bildoperatoren.....	37
2.4 Merkmale, Darstellungen und exemplarische Operationen.....	45
2.4.1 Merkmale, Kanten und Bereiche	45
2.4.2 Darstellungen.....	48
2.4.3 Exemplarische Operationen.....	53
3. Radiometrische und fotometrische Grundlagen der Bildentstehung	67
3.1 Radiometrische Begriffe	67
3.1.1 Punktförmiger Sender	69
3.1.2 Flächenhafter Sender	71
3.1.3 Empfänger	73
3.2 Fotometrische Begriffe.....	78
3.2.1 Punktförmiger oder flächenhafter Sender	79
3.2.2 Empfänger	80
3.3 Zusammenhang zwischen radiometrischen und fotome- trischen Größen	81
3.4 Farbdarstellungen der Drei-Komponenten-Theorie.....	90
3.4.1 Farbraum	92
3.4.2 RGB-Farbraum.....	93
3.4.3 XYZ-Farbraum.....	97
3.4.4 FHS-Farbraum.....	101

4. Lineare Kantenoperatoren und Filter	107
4.1 Kontinuierliche differentielle Operatoren.....	108
4.1.1 Richtungsabhängiger Kantenoperator Grauwertgradient.....	109
4.1.2 Richtungsunabhängiger Kantenoperator: Laplace-Operator.....	111
4.1.3 Richtungsunabhängiger Zwei-Komponenten-Operator: Laplace-Gauß-Operator	111
4.2 Digitale differentielle Kantenoperatoren.....	119
4.2.1 Vergleich durch Kreuzkorrelation	119
4.2.2 Ungemittelte digitale Gradienten.....	121
4.2.3 Gemittelte und gewichtete Gradienten	124
4.2.4 Digitaler Laplace-Operator	125
4.3 Gaborfilter.....	126
4.4 Eindimensionale rekursive Filter zweiter Ordnung.....	134
4.4.1 Canny'sche Gütekriterien für eindimensionale IIR- Kantenfilter.....	135
4.4.2 Eindimensionale Operatoren für Stufenkanten: Die Impulsantworten von Shen und Deriche.....	139
4.4.3 Rekursive Implementierung des eindimensionalen Kantenoperators von Deriche	145
4.4.4 Konstruktion zweidimensionaler IIR-Kantenoperatoren	147
4.5 Polynominterpolation zur Kantenextraktion und topo- grafische Pixelklassifikation.....	151
4.5.1 Interpolation durch Polynome erster Ordnung	151
4.5.2 Interpolation durch Polynome zweiter Ordnung.....	155
4.5.3 Statistische Tests bei linearer Polynominterpolation.....	159
5. Bildsegmentierung und Merkmalsextraktion	165
5.1 Bildsegmentierung	165
5.2 Repräsentation.....	166
5.3 Schwellwertverfahren	167
5.3.1 Einfaches Schwellwertverfahren.....	167
5.3.2 Multischwellwertverfahren	169
5.3.3 Variable Schwellwerte.....	169
5.4 Mehrkanalige Bilder	169
5.4.1 Anwendungsbeispiel: Luftbilder	170
5.4.2 Kontrast als Segmentierungskriterium.....	174
5.5 Flächenwachstumsverfahren.....	174
5.6 Texturanalyse.....	175
5.6.1 Verfahren von Laws.....	176
5.6.2 Beispiel: Gaborfilter.....	177

Vorwort der Herausgeber

Das *Handbuch der Informatik* versucht, das Gesamtgebiet der Informatik mit seinen Grundlagen, seinen Teilgebieten und seinen wichtigsten Anwendungen zusammenhängend darzustellen. Es informiert Lehrende und Lernende, DV-Praktiker und DV-Nutzer über Konzepte, Methoden und Techniken, deren grundlegende Bedeutung anerkannt ist, deren Nützlichkeit in der Praxis sich gezeigt hat; Grenzen und Entwicklungstendenzen eines Gebietes werden angesprochen.

Jeder Band des *Handbuch der Informatik* behandelt ein in sich abgeschlossenes Thema. Der Leser kann sich - unabhängig von den anderen Bänden des Handbuches - in das betreffende Gebiet neu einarbeiten, vorhandenes Wissen auffrischen oder im Sinne eines Nachschlagewerkes einzelne Themen aufspüren und vertiefen. Hierbei helfen Strukturierung und Typographie des Textes und die Hinweise auf weiterführende Literatur.

Es ist in letzter Zeit Mode, etwas süffisant zu fragen, "was bei der künstlichen Intelligenz eigentlich herausgekommen" sei. Dabei gibt es schon einige praxisrelevante Gebiete, auf die man solche impertinenten Frager hinweisen kann. Aber wohl keines eignet sich dabei so gut, "Punkte zu machen", wie die Bildverarbeitung.

Hier ist wirklich ein signifikanter Unterschied zu erkennen zwischen dem, was vor vielleicht 10 Jahren ein guter "klassischer" Informatiker konnte, und dem, was die Autoren dieses Buches uns heute als unmittelbar anwendbare Theorie und Technik erklären. Und niemand kann sagen, diese Künste seien brotlos. Neben der Bilderkennung, welcher das ursprüngliche Interesse der meisten KI-Spezialisten galt, nutzen inzwischen viele Anwender die Ergebnisse und Produkte, die auf modernen Bildbearbeitungsmethoden beruhen. Seien es so "primitive" Aufgaben wie die Qualitätsverbesserung und das Zoomen gescannter oder vom Bildschirm abgenommener Bitmaps oder - das andere Extrem - die Wunder, welche Computer-Graphiker uns täglich in Zeitschriften, auf dem Fernsehbildschirm und der Kinoleinwand vorführen.

Wir freuen uns deshalb besonders, Ihnen jetzt auch diese Schlüsseltechnologie der modernen Softwareentwicklung in unserem Handbuch kompetent präsentieren zu können.

A. Endres

H. Krallmann

P. Schnupp

Danksagung

Ein Handbuch über Bildverarbeitung zusammenzustellen, bedarf einer Zusammenarbeit vieler Beteiligter. Die Mitautoren haben in der Zeit, in der das Buch entstand, mit mir gemeinsam am Institut für Informatik an vielen Vorhaben aus dem Bereich des Verstehens von Bildern gearbeitet. Ergebnisse daraus und aus der Lehrtätigkeit in diesem Gebiet sind in das Buch eingeflossen. Die Zusammenarbeit im Autorenteam war angenehm und hat uns selbst in den Diskussionen neue Einsichten vermittelt.

Ohne Unterstützung hätten wir alle die Arbeit nicht fertig stellen können. Wir sind zu Dank verpflichtet

Herrn Werner Meixner für seine Diskussionsbeiträge und kritischen Kommentare,

Herrn Stefan Lanser für häufige Diskussionen und die Überlassung seiner Diplomarbeit,

Frau Susanne Gerl für die Überlassung ihrer Diplomarbeit,

Herrn Georg Schmid für die Überlassung seiner Seminararbeit,

Herrn Stefan Brüess für die Unterstützung bei der Anfertigung der Druckvorlagen,

Frau Evelyn Gemkow, Herrn Peter Bendel und Herrn Tobias Blank für die Erfassung und Formatierung des Textes,

Frau Adriana Faber für ihre Geduld und Bereitschaft, drei Kapitel vorlagefertig zu schreiben,

Frau Inge Zeller und Herrn Jens Wunderwald für die Durchsicht einiger Kapitel.

Von den Mitarbeitern des R. Oldenbourg Verlages, die unsere Arbeit begleitet haben, möchten wir besonders Frau Eva-Maria Schulz, Frau Susanne Spitzer und Frau Margarete Metzger erwähnen.

Wir wünschen den Lesern eine angenehme Lektüre und erfolgreich verwertbare Erkenntnisse. Im Namen der Autoren

Bernd Radig

1. Einleitung

Bernd Radig

Kenntnisse aus den Bereichen Sensortechnik, digitale Signalverarbeitung, Mustererkennung, Künstliche Intelligenz und anderen müssen zusammenkommen, um Programmsysteme für Computer zu entwerfen, die automatisch Bilder interpretieren, sie in gewissem Sinne verstehen. Wir beschränken uns in diesem Buch auf die Sichtweise, daß ein Bild von einem Programmsystem verstanden wurde, wenn es als Ergebnis jeweils einem Ausschnitt des Eingangsbildes genau einen Namen aus einer Liste von Begriffen zugeordnet hat. Das Bild wurde teilweise verstanden, wenn die mit Namen belegten Ausschnitte das Bild nicht vollständig überdecken. Es wurde richtig verstanden, wenn ein Bewerter aus der Namensbelegung korrekte Konsequenzen ableiten kann. Die Abbildung 1.0-1 zeigt einen Werkstattwagen, aufgenommen in einer Büroumgebung. Die Abbildung 1.0-2 stellt die Namensbelegung von Ausschnitten dar, wie sie ein typisches Interpretationsprogramm [Munkelt, Kristen 92] herstellt.



Abb. 1.0-1: Originalbild der Büroszene mit Werkzeugwagen



Abb.1.0-2: Interpretationsergebnisse für den Werkzeugwagen

In dem Bild sind einige Ausschnitte, allerdings nicht alle, mit denselben Namen belegt, die ein Betrachter verwenden würde. Ist das Bild korrekt interpretiert? Die Antwort hängt von der Weiterverwendung der Interpretationsergebnisse ab. Oftmals ist ein bildverstehendes

System Bestandteil einer umfassenden Problemlösung, in der die Korrektheit des Interpretationsergebnisses durch den Erfolg definiert wird, mit dem das umgebende System die Ergebnisse verwendet.

Der Leser versteht diesen Text, der vor ihm liegt, weil er den hellen Ausschnitten dieser Buchseite den Namen "Papier" gibt und jedem individuellen schwarzen Ausschnitt den Namen des betreffenden Buchstabens zuordnet. Das Ergebnis ist im Sinne des Bildverstehens korrekt, wenn der Leser aus den getroffenen Zuordnungen die vom Autor intendierten Wörter und Sätze rekonstruieren kann. Ob diese inhaltlich verständlich oder korrekt sind, hat allerdings nichts mehr mit Bildverstehen zu tun. Um die Aufgabe, Bilder automatisch zu interpretieren, etwas präziser zu beschreiben, haben wir der Begriffsdefinition des Bildverstehens ein eigenes Kapitel gewidmet.

Wir haben in den darauf folgenden Kapiteln den Weg gewählt, den Prozeß des Bildverstehens durch alle Schritte der Abstraktion zu verfolgen, von den radiometrischen und fotometrischen Grundlagen über die Ebene der symbolischen Beschreibung bis hin zur Belegung mit Namen durch Zuordnung von Modellen der benannten Gegenstände zu in Bildausschnitten vorgefundenen Strukturen. Hinweise zur Architektur bildverstehender Systeme und auf beispielhafte Problemlösungen runden die Darstellung ab.

Um dieses Ziel unter Beschränkung auf den bereitgestellten Platz zu erreichen, mußten wir vieles unerwähnt lassen. Wir sind nicht auf die ganze Bandbreite der heute zur Verfügung stehenden bildgebenden Sensoren eingegangen, sondern haben die Grundlagen nur für den visuellen Bereich behandelt. Dabei haben wir als übliches Bildformat ein rechteckiges Gitter mit $n \times m$ Zeilen und Spalten angenommen. Sensoren, die eine andere Form der künstlichen Retina implementieren, solche mit anderer Topologie, ortsvarianter Auflösung, integrierter Vorverarbeitung oder weiteren Besonderheiten werden erst in Zukunft Bedeutung erlangen.

Hinweise auf Ultraschallbild-Interpretation fehlen beispielsweise genauso wie eine gründliche Darstellung der Röntgen- oder MR-Tomographie. Dafür haben wir einen breiteren Raum der Behandlung von Farbe gewidmet in dem Bewußtsein, daß technologische Fortschritte das Ausnutzen der Farbinformation erlauben, ohne den Interpretationsprozeß dadurch unverhältnismäßig zu verlängern, daß ein wesentlich größeres Datenvolumen gegenüber Grautonbildern verarbeitet werden muß. Hier wird sich die analoge Entwicklung vollziehen wie in der industriellen Bildinterpretation vor einiger Zeit beim Übergang vom Binärbild zu Grautonbildern.

Wir haben das weite Feld der Deutung von Bildfolgen ausgelassen, obwohl uns bewußt ist, daß die Interpretation von Einzelbildern, wenn sie Schnappschüsse einer sich dynamisch verändernden Umgebung sind, durch die fehlende Information darüber erschwert wird, wie die auf-

genommenen Gegenstände sich bewegen oder ihr Erscheinungsbild ändern.

In den letzten dreißig Jahren ist der Fundus an Verfahren, die in den einzelnen Schritten des automatischen Bildverstehens eingesetzt werden können, so gewaltig angewachsen, daß wir nur noch eine sehr schmale und subjektive Auswahl treffen konnten. Sie stellt jedoch die Theorie der linearen Operatoren und Filter einigermaßen vollständig dar und ist im Bereich der Bildsegmentierung und Merkmalsextraktion repräsentativ.

Für die Vorgänge auf der Ebene der symbolischen Beschreibung der Bilder und den darauf aufbauenden Interpretationsverfahren konnten wir nur noch beispielhaft eine mögliche Vorgehensweise und Methodik beschreiben. Die wesentlichen Aufgaben der Interpretationsschicht sind mit exemplarischen Lösungen dargestellt. Entwurfsentscheidungen, die hier zu treffen sind, umfassen unter anderem

- die formale Repräsentation der Bildbeschreibung,
- die Repräsentation der Modelle, die prototypisch Gegenstände mit ihren verwertbaren Eigenschaften, ihren möglichen Anordnungen, Beleuchtungsverhältnissen, Sensoreigenschaften und im weitesten Sinne den Interpretationskontext beschreiben,
- die Verfahren, mit denen Bildstrukturen und Modellstrukturen verglichen werden,
- die Definition des Begriffes Ähnlichkeit, formaler ausgedrückt, einer Ordnungsrelation zwischen möglichen Zuordnungen zwischen Bild- und Modellstrukturen,
- Methoden des Schließens, um aus erfolgreichen und mißlungenen Zuordnungen von Teilstrukturen Interpretationshypothesen oder Verbesserungsversuche abzuleiten,
- Adaption von Modellstrukturen an aus dem Bild gewonnenen Erkenntnissen, beispielsweise wenn es Hinweise auf Verdeckungssituationen gibt, bei denen eine Ähnlichkeit mit dem Modell und dem verdeckten Teil des Gegenstandes überhaupt nicht zu erwarten ist,
- Definition von akzeptablen Zuordnungen und von Heuristiken, die den Suchprozeß nach besten Zuordnungshypothesen beenden und den Suchraum einschränken.

Wir haben uns bemüht, eine im Buch einheitliche Begriffswelt zu definieren und mit angemessenem Aufwand zu präzisieren. Für das Thema Bildverstehen brauchen wir Begriffe für Elemente aus drei Bereichen. Der erste ist die reale Welt, die Umgebung des Sensors, aus der Bilder aufgenommen und interpretiert werden sollen. Begriffe dort bezeichnen Gegenstände, Ereignisse, Beziehungen zwischen Dingen, Vorgänge und viele weitere Phänomene. Der zweite Bereich umfaßt die Bilder, deren Inhalt auf unterschiedlichen Abstraktionsebenen durch Verarbeiten und Weiterverarbeiten des Sensorsignals repräsentiert

wird. Hier sind beispielsweise Begriffe zu bilden, die Eigenschaften des Sensors charakterisieren, Merkmale abstrakter Einheiten bezeichnen oder Beziehungen zwischen Einheiten benennen. Der dritte Bereich umfaßt die Welt der Modelle. Sie lenken die Bildinterpretation durch Vorgabe der Abstraktionshierarchie und Bereitstellung – in geeigneter Repräsentationsform – des Wissens, das der Interpretationsvorgang benötigt.

Begriffe wie Objekt, Struktur, Merkmal, Attribut, Primitiv, Komponente und viele weitere werden in den drei Bereichen oft ohne Präzisierung und Angabe ihrer dort intendierten Bedeutung gebraucht. Nachlässigkeiten in der Sprache des Buches haben wir auch deswegen zu meiden versucht, um durch irreführende Wortassoziationen gar nicht erst Aussagen zu provozieren wie »Das System hat den Verkehrspolizisten erkannt« oder »Der Computer kann sehen«. Obwohl wir selbst verkürzte Aussagen dieser Art gelegentlich gebrauchen, haben wir sie in einen solchen Kontext gestellt, daß ihre präzise Bedeutung erschließbar wird.

Verstehen heißt benennen, also Feststellen von Klassenzugehörigkeit der dann mit dem Klassennamen belegten Bildausschnitte. Klassifikation kann auf unterschiedlichen Abstraktionsebenen stattfinden, signalnah beispielsweise durch Approximation der ersten Ableitung der Bildfunktion und Selektion der lokalen Extrema als Kantenpunkte. Das Bilden von Merkmalsvektoren erlaubt den Rückgriff auf Verfahren der Mustererkennung zur Klassifikation. Müssen Bildausschnitte durch einzelne Komponenten mit vielfältigen, gegenseitigen Beziehungen beschrieben werden, so werden Verfahren zum Vergleich bildbeschreibender Strukturen mit modellbeschreibenden Strukturen eingesetzt, die auch iterativ auf einigen Abstraktionsebenen angewendet werden können.

Geht das Ziel des Bildverstehens über ein einfaches Bezeichnen von Bildausschnitten hinaus, weil komplexe Strukturen, Zusammenhänge oder Ereignisse beschrieben werden sollen, so werden die Verfahren des Strukturvergleichs zwischen Modell und Bild zunehmend durch Methoden aus dem Repertoire der Künstlichen Intelligenz ergänzt und ersetzt. Weitere Aufgaben können dann bewältigt werden wie etwa das automatische Lernen von systeminternen Modellen oder zumindest deren Anpassung an die aktuellen Beobachtungen.

Gleichzeitig eröffnet sich die Möglichkeit, eine weitergehende Interpretation des beobachteten Bildes nahtlos anzuschließen, um nicht nur die Frage zu beantworten, was die Kamera sieht sondern auch, warum sie etwas sieht oder auch nicht. Die Kamera eines autonomen Fahrzeuges liefert auch mit nachgeschalteter modellbasierter Interpretation allein noch nicht genügend Information, um das Fahrzeug vernünftig zu steuern. Erst wenn der Interpretationsvorgang der einzelnen

Bilder, die während der Fahrt aufgenommen werden, durch Folgerungen über Gefährdungssituationen und vergangenes sowie zu erwartendes Geschehen ergänzt wird, ist eine optimierte Steuerung des Fahrzeugs möglich.

In manchen Anwendungen sind ja sogar Aussagen darüber, was nicht zu sehen ist, wichtiger als die Klassifikation von Bildausschnitten, etwa wenn eine Palette mit Ladegut für das autonome Fahrzeug nicht am erwarteten Platz steht. Daß auf dieser Ebene des Schließens über die Bedeutung von Bildausschnitten elegant Informationen von weiteren Sensoren, die in einer industriellen Anwendungsumgebung häufig zu finden sind, integriert werden können, sei nur nebenbei bemerkt. Beispielsweise um in einer Fabrikumgebung das Hindernis im Weg eines Fahrzeugs zu analysieren, läßt sich wirkungsvoll der Doppler-Effekt eines Radar-Sensors ausnutzen, um zwischen stationärem Gegenstand, Fahrzeug oder Mensch zu unterscheiden. Die Unterscheidung zwischen beiden letzteren ist möglich, weil sich der Mensch viel ungleichförmiger als ein Fahrzeug bewegt und sogar im Stehen Körperschwingen nicht unterdrücken kann.

Die folgenden Kapitel bieten Unterstützung für die ganze Spannweite des Prozesses an, Bilder automatisch zu verstehen. Wir haben für jede Abstraktionsebene des Prozesses Beispiele gegeben – von den radiometrischen Grundlagen bis hin zu modellbasierten Interpretationsverfahren. Wir hoffen, daß auf dieser Basis aufbauend dem Leser die Erschließung aktueller Fragestellungen leichter fällt. Weitgehend unbefriedigend gelöste Probleme bestehen noch in vielen Bereichen, beispielsweise

- Rekonstruktion der Tiefeninformation aus Einzelbildern, Stereobildern oder Bildfolgen sowie die Integration von Methoden aus verschiedenen Ansätzen. Einzelne Verfahren versagen in ungeeigneten Bildbereichen, etwa weil sie durch Rauschen besonders stark gestört werden, lokal ausreichend steile Intensitätsübergänge benötigen oder eine Textur auswerten, die unter flachem Blickwinkel nicht mehr aufgelöst werden kann. Kombinationen von jeweils zwei Verfahren, etwa Analyse des Intensitätsverlaufes gekoppelt mit dem Auswerten von Stereobildern oder Folgen aus Stereo-Bildpaaren, werden seit einigen Jahren systematisch untersucht.
- Interpretationen können dadurch verbessert werden, daß Information aus mehreren Bildern kombiniert wird. Das gleichzeitige Ausnutzen von Entfernungs- und Intensitätsinformation, geliefert von Laser-Kameras, ist ein schon klassisches Beispiel, Integration der Information aus mehreren spektralen Kanälen bei der Fernerkundung ein anderes. Das automatische Zusammensetzen eines Bildes mit großem Tiefenschärfenbereich aus einzelnen, unterschiedlich fokussierten

Bildern ist ein ganz neues Verfahren, dessen Einsatzmöglichkeiten noch nicht ausgelotet sind.

- Vielen Verfahren der Objektidentifikation und Formrekonstruktion liegen vereinfachte Annahmen der Reflexionseigenschaften der Objektoberfläche und der Beleuchtung zu Grunde. Hinzunahme von Farbinformation verspricht einen Gewinn an Robustheit und eine Erweiterung des Einsatzspektrums. Neuere Beleuchtungsmodelle erlauben die Definition und Positionierung mehrerer Punktlichtquellen, das Einbeziehen von Sekundärstrahlung und diffusem Licht. Die Schätzung einer Oberflächennormalen bei diffusem Licht erfordert jedoch ganz andere Ansätze als bei Punktlicht-Beleuchtung. Neuere Arbeiten beschäftigen sich mit dem Zusammenhang zwischen Reflektanz, Beleuchtung und Oberflächenneigung mit der Absicht, nicht mehr allein die beobachteten Oberflächen im dreidimensionalen Raum zu rekonstruieren, sondern auch Reflektanzeigenschaften oder Zahl und Position von Lichtquellen zu ermitteln.
- Gerade der modellbasierte Ansatz zur Interpretation von Bildern ist auf die Qualität der verwendeten Modelle angewiesen. Detaillierte Modelle für in der Realität vorkommende Gegenstände "in Handarbeit" zu erzeugen, ist mit erheblichem Aufwand verbunden. Ein Ausweg bei gefertigten Objekten ist die Verwendung der dann oftmals vorliegend CAD-Konstruktionsunterlagen für die Modellbildung, in erster Linie für Geometriedaten. Sollen jedoch Gegenstände beschrieben werden, die beliebig gewölbte Oberflächen besitzen, einem generischen Bauplan folgen, wie bei Pflanzen, aber bei jedem individuellen Exemplar variieren oder verformbar sind und auch ihre Form laufend ändern, wie bei Lebewesen, so ist die Wahl einer kompakten dreidimensionalen Repräsentation genauso schwierig wie der Aufbau eines konkreten Modells. Ansätze, solche "schwierigen" Gegenstände unter kontrollierten Bedingungen zu beobachten, durch Tiefenmessungen abzutasten oder durch Invarianten, beispielsweise automatisch ermittelte Symmetrieeigenschaften, zu beschreiben, werden schon seit mehr als zwanzig Jahren publiziert, gehören aber noch nicht zum gesicherten Fundus für den Aufbau bildverstehender Systeme.

Mit dieser Aufzählung ist der Bereich interessanter und herausfordernder Fragestellungen längst nicht ausgeschöpft. Fortschritte in den einzelnen Teilbereichen bilden die Basis für den Aufbau leistungsfähiger und flexibler Systeme. Können Verfahren, die zum automatischen Verstehen von Bildern beitragen, in aktiv sehende Systeme integriert werden, so eröffnen sich neue Möglichkeiten, die beobachtete Umwelt zu analysieren. Die zielgerichtete Variation des Beobachtungsstandortes und der Beleuchtung durch Kamera und Scheinwerfer tragende Manipulatoren oder Fahrzeuge läßt es beispielsweise zu,

störende Glanzlicht-Reflexionen zu vermeiden, Verdeckungen teilweise oder vollständig zu beseitigen, Objekt- und Umgebungsmodelle durch systematische Beobachtung zu lernen oder den Fokus der Analyse, gesteuert durch das Ergebnis vorangegangener Interpretationen, auf interessante Ausschnitte der Umgebung zu konzentrieren.

Der geplante Umfang des Buches läßt Ausflüge in diese aufregenden Gebiete nicht zu. Einen Einblick in die aktuellen Forschungsergebnisse geben neben den einschlägigen Zeitschriften unter anderem die Tagungsbände des Symposiums der Deutschen Arbeitsgemeinschaft für Mustererkennung, der European bzw. International Conference on Computer Vision oder der International Conference on Pattern Recognition.

Wir haben uns in diesem Buch allerdings nicht nur darauf beschränkt, Grundlagen und Methoden für die einzelnen Phasen des Bildverstehens aufzuzählen, sondern haben ihre Integration in ein Bildanalyse-Werkzeug genauso beschrieben wie ihre Nutzung in beispielhaften Anwendungen.

Insbesondere in den Kapiteln über die Architektur von Bildverarbeitungssystemen und die beispielhaften Problemlösungen taucht ein neuer und noch längst nicht ausreichend behandelter Fokus des Buches auf, nämlich die Frage, wie Problemlösungen systematisch, effizient, reproduzierbar und wiederverwendbar konstruiert werden können. Das ingenieurmäßige Vorgehen dazu wird hier auf einem recht abstrakten Niveau behandelt. Die Abbildung auf heute industriell eingesetzte Hardware- und Software-Systeme mußten wir allerdings genauso stiefmütterlich behandeln wie die Unterstützung des Problemlösungsprozesses durch automatisierbare Verfahren. Dazu ist auch das Feld des »Computer Aided Vision Engineering« noch zu jung.

2. Grundlagen und Begriffsdefinitionen des Bildverstehens

Paul Levi

Bildverstehen (Bildinterpretation, Maschinelles Sehen, Computersehen) wird in Analogie zum menschlichen Sehen definiert. Man versteht daher unter *Bildverstehen* einen mehrschichtigen Prozeß, bei dem es darum geht, Bilder zu digitalisieren und so aufzubereiten, daß sie durch den Vergleich mit zuvor gespeicherten Modellen (z.B. Objekt- und Szenenmodelle) rechnergestützt *vorverarbeitet*, *erkannt* (identifiziert) und letztlich *verstanden* (interpretiert) werden können [Binford 82], [Neumann 85]. Der Begriff des Verstehens umfaßt ganz allgemein die drei folgenden, verschiedenen Bedeutungen:

Bildverstehen

- 1) Bedeutungserfassung einer Bildstruktur durch die Zuweisung eines Namens.
- 2) Begreifen von ursächlichen Zusammenhängen (kausales Verstehen).
- 3) Intuitives Erfassen von Zusammenhängen (intuitives Verstehen) im Gegensatz zum kausalen Verstehen.

In der Bildverarbeitung meint man unter Verstehen primär die Zuweisung eines eindeutigen Namens, der eine feste Bedeutung (z.B. Tisch, Stuhl, Gesicht) hat, an eine Bildstruktur (wird später definiert). Durch die Namenszuweisung verstehen wir, was ein Bildstruktur ist. Die zweite und dritte Bedeutung des Verstehens liegen gegenwärtig noch weit außerhalb der Reichweite des jetzigen Bildverstehens.

Die Mehrschichtigkeit dieses Prozesses spiegelt sich wider in den verschiedenen Abstraktionsebenen der verwendeten Modelle (z.B. Bildmodell, Komponentenmodell, Szenenmodell), in den eingesetzten Operationen (Vorverarbeitung, Segmentierung, Komponentenerkennung, Objekterkennung, Szenenerkennung) und in der Steuerung dieser Operationen (top-down, bottom-up oder kombiniert). Unter einer *Operation* verstehen wir dabei eine Funktion, die von einer Darstellung in eine andere Darstellung (z.B. Bildmerkmal in Merkmalsmodell, relational dargestellte Bildkomponente in entsprechendes Komponentenmodell) abbildet.

Man unterteilt das Bildverstehen in mindestens vier Schichten, wobei die Namensgebung dieser Schichten durch die typischen, darin vorkommenden Operationen bestimmt wird (Abbildung 2.1-1). Bildverstehen beginnt ab der dritten Schicht, in der die aus dem Bild erzeugte symbolische Objektbeschreibung mit symbolischen Objektmodellen verglichen wird. Bildverstehen wird häufig auch als Szenenanalyse be-

Operations-
klassen

zeichnet [Kazmierczak 80]. Die nächsttiefere Bildverarbeitungsschicht bildet die Komponentenerkennung. Sie vergleicht die aus dem Bild erzeugten Bildkomponenten mit Komponentenmodellen. Ganz unten ist die Bildvorverarbeitung angesiedelt, die die Bildverarbeitung im engeren Sinne (Bildsignalverarbeitung, Bild zu Bild Transformation) darstellt.

In diesem Kapitel bauen wir zuerst ein elementares Schichtenmodell des Bildverstehens auf, um uns anhand dieses Modelles bei der Bildung von Begriffsdefinitionen zu orientieren. Diese Definitionen konzentrieren sich vor allem auf die Modelle und Operationen, die zur Komponenteklassifikation und Objekterkennung führen. Die im Teilabschnitt 2.1 verwendeten Begriffe werden bei Bedarf in den nachfolgenden Teilabschnitten weiter präzisiert.

2.1 Schichtenmodell des Bildverstehens

2.1.1 Deklarative Wissens Ebenen des Bildverstehens

Modelle

Aus der Sicht der Informatik ist Bildverstehen ein Vorgang, bei dem zuerst Modelle aufgebaut werden müssen, um zu definieren, was überhaupt extrahiert (gemessen) und gesehen werden kann. Dabei verstehen wir unter einem *Modell* die Darstellung sowohl prototypischer Eigenschaften und gegebenenfalls der Strukturen einer Klasse von Bildprimitiven, Bildkomponenten, Bildobjekten oder Szenen als auch eines idealen Individuums dieser Klassen. Zur internen Unterscheidung teilen wir die Modelle in *Klassen-* und *Individualmodelle* ein, wobei Klassenmodelle die Eigenschaften mehrerer Modelle abstrahierend zusammenfassen, während die Individualmodelle nur ein einziges Modell enthalten. Die Klassenmodelle werden benutzt, um den häufig gewollten Vergleich zwischen Individualmodell und dem entsprechenden Bildelement (z.B. Bildkomponente) vorzubereiten, denn aus der Kenntnis der Klassenzugehörigkeit läßt sich die Zuordnung eines Bildelementes zu dem korrespondierenden Individualmodell häufig zuverlässiger durchführen. Ein Bildelement charakterisiert dabei alles, was in einem Bild, unabhängig von der gewählten Darstellung, vorhanden sein kann und wird aus der Aufnahme eines physikalischen Objektes erzeugt. Wir verwenden den Begriff eines Bildelementes daher typischerweise für Pixel, Bildprimitive, Bildkomponenten, Bildobjekte und Bildszenen. Die Realisierung eines Klassen- oder Individualmodelles durch ein bestimmtes Bildelement wird häufig auch als *Exemplar* ("Instanz") dieses Modelles bezeichnet. In manchen Anwendungen wie der Musterklassifizierung (z.B. Schrift- und Spracherkennung) verwendet man lediglich Klassenmodelle, da es nur darum geht, ein Bildelement einer Klasse zuzuordnen (zu klassifizieren). Ferner können Modelle durch Verallgemeinerungs- oder Spezialisierungsrelationen zum Beispiel mit semantischen Netzen hierarchisch aufgebaut werden (Abschnitt 2.4.2).

Die Begriffe eines Objektes und einer Bildstruktur werden sehr unterschiedlich gebraucht. Wir verwenden diese Begriffe in den folgenden Bedeutungen. Ein *Objekt* ist ein (meist strukturierter) realer Gegenstand und nicht im Sinne der objektorientierten Programmierung als beliebige Einheit zu verstehen. Eine *Bildstruktur* ist eine aus Bildprimitiven oder Bildkomponenten zusammengesetzte Struktur, die auch selbst wieder Teil komplexer Bildelemente (z.B. relational dargestellte Bildszene) sein kann. Eine Bildstruktur ist daher eine relationale Darstellung, die z.B. die jeweiligen wechselseitigen Beziehungen von Bildprimitiven (Segmentierung) oder von Bildkomponenten (Komponentenerkennung) beschreibt. Formal läßt sich der Begriff einer Bildstruktur zwar auch auf relationale Darstellungen von Pixelmengen (Bildvorverarbeitung) anwenden, doch verbinden wir damit keine relevante Bedeutung und verwenden diesen Begriff daher nicht auf der Ebene der Bildvorverarbeitung.

Bildstruktur

Merkmale sind charakteristische Eigenschaften (Attribute) einer Menge von Bildpunkten (Vorverarbeitung), von Bildprimitiven (Segmentierung), Komponenten oder von Bildstrukturen. Jede Hierarchieschicht des Bildverstehens hat ihre eigenen Merkmale. Beispiele sind statistische Mittelwerte, Schwerpunkte, Momente und Formmerkmale wie Kreisförmigkeit, Zylinderförmigkeit, Kollinerität usw. Merkmale sollten einen so hohen Informationsgehalt haben, daß mit ihrer Hilfe Bildelemente klassifiziert oder identifiziert werden können. Eine Merkmalsextraktion entspricht einer Messung. Die im Bild extrahierten Merkmale nennen wir Bildmerkmale und die entsprechenden Merkmale im Modell heißen Modellmerkmale.

Merkmal

Der Begriff eines Merkmales wird jedoch nicht immer in diesem präzisen Sinne verwendet. Viele Autoren bezeichnen das Bildelement selbst (z.B. Eckpunkt, Kante, Bildkomponente) als ein Merkmal und sie benutzen den Begriff erst nach der Bildvorverarbeitung, d.h. ab der Komponentenebene. Wir unterscheiden zwischen dem Merkmalsträger selbst (z.B. Kreis) und seinen Eigenschaften (z.B. Kreisförmigkeit).

Die Operationen auf den Bildelementen und Modellen unterteilen wir in die beiden Hauptklassen *Extrahieren* und *Vergleichen*. Es sind Merkmale oder Bildstrukturen, die extrahiert und verglichen werden. Diese prinzipielle Zweiteilung der Operationsklassen entspricht dem üblichen Ansatz in der Signalverarbeitung. Die Extraktionsoperationen entsprechen den Messungen und die Vergleichsoperationen sind den Meßdatenauswertungen gleichzusetzen. So entspricht eine Merkmalsextraktion auf der Komponentenebene einer Messung im vorverarbeiteten Bild. Die darauf aufbauende Erkennung von Komponenten mit Hilfe von Merkmalen ist analog zu einer Meßdatenauswertung. Den Extraktionsoperationen der ersten Schicht gehen vorverarbeitende Operationen voraus. Dieser Prozeß des Vorverarbeitens dient dazu, ein Eingabebild so zu verändern (vorzubereiten), daß aus ihm überhaupt Merkmale zu-

Extrahieren und
Vergleichen

2. Grundlagen und Begriffsdefinitionen des Bildverstehens

verlässig extrahiert werden können, die danach mit den entsprechenden Modellen verglichen werden. Die Vergleichsoperationen bilden in jeder Schicht die zuvor extrahierten Bildelemente (z.B. Komponenten) in die jeweiligen Modelle ab.

Erwarten

Die Operationen des Extrahierens und Vergleichens können jeweils durch entsprechende Erwartungs- bzw. Vorhersageoperationen unterstützt werden. Als Erwartungen bezeichnen wir die Einschränkung möglicher Modelle auf ausgewählte Modelle, um die Komplexität des Lösungsvorganges zu reduzieren. Im Bereich der Bildvorverarbeitung nennt man die Erwartungen Vorhersagen.

Klassifizieren und Vergleichen

Die Vergleichsoperationen bilden von den Bildelementen (z.B. Bild, Bildstruktur) auf die entsprechenden Modelle (Klassenmodelle, Individualmodelle) ab. Eine typische Vergleichsoperation ist eine Funktion, die ein Bildmerkmal auf ein Modellmerkmal abbildet, um das Bildmerkmal zu erkennen. Wir sprechen nach Niemann von einer *Klassifizierung* [Niemann 83], wenn die Vergleichsoperation die Zugehörigkeit eines Bildelementes zu einem Klassenmodell bestimmt und wir nennen die Korrespondenz zwischen einem Bildmodell und einem Individualmodell eine *Erkennung* im Sinne einer Identifikation.

Wissensbasiertes Bildverstehen

Legt man für ein Bildverarbeitungssystem eine wissensbasierte Architektur zugrunde, d.h. nimmt man eine Dreiteilung in eine Datenbasis, eine Wissensbasis und eine Inferenzmaschine (Kontrollkomponente) vor, dann muß zwischen deklarativem und prozeduralem Wissen genau unterschieden werden [Ballard, Brown 82]. Insgesamt wird *wissensbasiertes* Bildverstehen vor allem dadurch charakterisiert, daß zum einen diese beiden Wissensarten *explizit* implementiert worden sind und daß zum anderen Operatoranwendungen als *Suchvorgänge* und nicht als numerische Lösungsverfahren betrachtet werden. Reale wissensbasierte Sichtsysteme enthalten jedoch auch klassische, numerische Operationen (z.B. Bildvorverarbeitung), wobei aber der interne Aufbau und die Operatoren des Systems wissensbasiert sind und den nicht wissensbasierten Anteil dominieren [Liedtke, Ender 89].

Deklaratives Wissen

Als *deklaratives* Wissen bezeichnen wir alle Darstellungen, die keine Angaben darüber, wie diese Beschreibungen abgearbeitet sind (Ablaufsteuerung), enthalten. Daher werden Modelle und die Beschreibungen der zu den Modellen gehörenden Operationen von uns als deklarativ bezeichnet. Eine Menge von Produktionsregeln (Expertensystem) oder von Ersetzungsregeln (Grammatik) stellt somit deklaratives Wissen dar. Das Wissen über die explizite Steuerung des Operationsablaufes schließen wir davon aus, sie gehört zur prozeduralen Wissensart. Zur Verdeutlichung stelle man sich das Integrationskalkül (Differentialkalkül) vor. Hier entsprechen die Modelle den einzelnen Integraltypen wie Grundintegral (z.B. Potenzen, Hyperbelfunktionen, rationale bzw. irrationale Funktionen), Kurvenintegral, Flächenintegral, räumliches Integral, usw. Mögliche Operationen darauf sind Substitutionen

(algebraische, trigonometrische), Partialbruchzerlegungen, quadratische Ergänzungen, partielles Integrieren, usw. Die Festlegung, welcher Integraltyp durch ein bestimmtes Integral (Problemdefinition) gegeben ist und welche Operation (z.B. partielles Integrieren) auf dieses Integral anzuwenden ist, wird mit Hilfe von prozeduralem Wissen durchgeführt.

Unter *prozeduralem* Wissen fassen wir das Wissen über die globale Steuerung des gesamten Systems (Inferenzmaschine), die Auswahl der geeigneten Operationen und die interne Steuerung (z.B. Parameteranpassung) dieser Operationen bei der Bildauswertung zusammen. Diese Wissensdefinitionen orientieren sich an den in der Programmierung üblichen. Hier unterscheidet man auch zwischen der Menge von unabhängigen Programmanweisungen (deklaratives Wissen), der Reihenfolge der Programmanweisungen und den Kontrollkonstrukten in diesem Programm. Diese festgelegte Reihenfolge, die Kontrollstrukturen und die von dem Betriebssystem verwendeten Kontrollanweisungen zur Programmabarbeitung (Prozeß) entsprechen dem prozeduralen Wissen. In diesem Teilabschnitt beschäftigen wir uns mit dem deklarativen und im nächsten mit dem prozeduralen Wissen.

Prozedurales
Wissen

Abbildung 2.1-1 zeigt das Schichtenmodell des Bildverstehens. Es verdeutlicht die hierarchische Organisation des dabei verwendeten deklarativen Wissens und läßt den Kontrollzyklus (prozedurales Wissen) vorerst außer acht. Dieser Ansatz stützt sich primär auf Modelle, die in vier Abstraktionsebenen hierarchisch angeordnet sind. Diese Modelle beschreiben durch Restriktionen, welche verallgemeinerten (generischen) Eigenschaften oder strukturellen Verbindungen erfüllt sein müssen, damit entweder die Klassenzugehörigkeit (Klassenmodell) eines Bildelementes (z.B. Komponente) oder die Individualität (Individualmodell) dieses Bildelementes mit Hilfe der zugehörigen Vergleichsoperationen (z.B. Schablonenvergleich oder relationaler Vergleich) bestimmt werden kann. So beschreiben Bildmodelle die Bedingungen, die von Pixeleigenschaften erfüllt sein müssen (z.B. Wahrscheinlichkeitsverteilungen, Leistungsspektrum, Begrenzungen von Ortsfrequenzen). Strukturelle Vorgaben sind auf Pixelebene nicht angebracht. Strukturelle Festlegungen werden erst dann notwendig, wenn Bildkomponenten, ihre Eigenschaften und ihre wechselseitigen Beziehung modelliert werden müssen. Diese relationale Modellierung ist ebenfalls für alle Ebenen des Bildverstehens wie Objekt- und Szenenbeschreibung unumgänglich.

Schichten-
modell

Die Modelle der einzelnen Schichten können prinzipiell sowohl für die Operationen der Extraktion (und des Erwartens) als auch für die Operationen des Vergleichens (und Erwartens) gebraucht werden. In der ersten Schicht allerdings werden die Bildmodelle hauptsächlich für die Vorverarbeitung und nicht für den Vergleich eingesetzt. Dies ändert sich ab der zweiten Schicht in wachsendem Maße hin zu Modellen (Komponentenmodell, Objektmodell, Szenenmodell), die vor allem für die Vergleichsoperationen eingesetzt werden. Innerhalb des gesamten

2. Grundlagen und Begriffsdefinitionen des Bildverstehens

Schichtenmodelles hat die Modellierung bis auf die Vorverarbeitung für die Vergleichsoperationen eine größere Bedeutung als für die Extraktionsoperationen, daher wurde der Einsatz von Modellen nur in der rechten Seite von Abbildung 2.1-1 explizit eingezeichnet.

	Extrahieren	Modelle	Vergleichen
4	$\text{Objekt}'_B \rightarrow \text{Szene}_B$	Szene_M	$\text{Szene}'_B: \text{Szene}_B \rightarrow \text{Szene}_M$
3	$\text{Komp}'_B \rightarrow \text{Objekt}_B$	Objekt_M	$\text{Objekt}'_B: \text{Objekt}_B \rightarrow \text{Objekt}_M$
2	$\text{Bild}'_B \rightarrow \text{Komp}_B$	Komp_M	$\text{Komp}'_B: \text{Komp}_B \rightarrow \text{Komp}_M$
1	$\text{Rohbild}_B \rightarrow \text{Bild}_B$	Bild_M	$\text{Bild}'_B: \text{Bild}_B \rightarrow \text{Bild}_M$

Abb. 2.1-1: Schichtenmodell des Bildverstehens. Der Index B steht für Bild und M soll auf Modell hinweisen.

Darstellungs-
ebenen

Die mathematischen Darstellungen, die man zur Beschreibung der Bildelemente und der Modelle benutzt, unterteilt man in vier Arten: Matrix, String (Zeichenfolge, Kette), relationale Darstellungen (Baum, Graph, Relationalstruktur, Abschnitt 2.4.2) und Grammatik. Die *Matrixebene* (Vektorebene) wird für Bildmodelle verwendet und beschreibt Grauwerte, Farbwerte oder verschiedene Leuchtdichten (Abschnitt 3.2) multispektraler Bilder. *String-Darstellungen* werden sowohl zur Codierung von Originalbildern als auch zur Beschreibung von Bereichen (z.B. Kettencodierung) in Komponentenmodellen verwendet. Diese String-Darstellungen sind jedoch nicht mit den korrespondierenden Darstellungen für String-Grammatiken zu verwechseln. Die einfachsten *relationalen Darstellungen* werden als Bäume in Komponentenmodellen zur zweidimensionalen Bereichsbeschreibung als Quadrees (quadratic trees) oder als Octrees (octal trees) zur dreidimensionalen Raumspezifikation verwendet. Ab der Komponentenebene werden aber auch Graphen und Relationalstrukturen als Darstellungen eingesetzt. Weitere Details zu diesen drei Darstellungsebenen sind im Teilabschnitt 2.4.2 zu finden. Als *Grammatiken* (syntaktische Darstellung) werden neben eindimensionalen String-Grammatiken zur Zeichenerkennung auch zweidimensionale Array- und Graphgrammatiken bzw. entsprechende attributierte Grammatiken eingesetzt. Der syntaktische Zugang zum Bildverstehen wird in diesem Buch nicht diskutiert, wir verweisen daher auf [Fu 77], [Niemann 81], [Bunke 85].

Nachdem wir die Modellbildung in unserem Schichtenmodell deutlich hervorgehoben haben, wollen wir darangehen, die einzelnen Schichten näher zu beschreiben.

Bildvorverarbeitung

In der ersten Schicht erfolgt die *Bildvorverarbeitung*. In ihr wird das Eingangsbild (Rohbild) zuerst digitalisiert (abgetastet und quantisiert) und danach kann es entzerrt (mit Sensormodell), komprimiert (z.B. Karhunen-Loève Transformation), verbessert (ohne zusätzliches Störmodell), restauriert (mit zusätzlichem Störmodell), rekonstruiert (z.B. mit zusätzlichem Gewebemodell für die medizinische Diagnose) und normiert (z.B. Bildgröße, Lage, Grauwertskalierung) werden.

Zu den extrahierenden (aufbereitenden) Operatoren zählen die Digitalisierung, Entzerrung, Komprimierung, Verbesserung und Normierung. Das Bild, das dadurch entsteht, wird als $Bild_B$ bezeichnet. Die Entzerrung eines digitalen Bildes dient dazu, geometrisch verzerrte Bilder, die typischerweise durch Sensorfehler (z.B. Kissenverzerrung durch eine Linse) zustandekommen, wieder zu korrigieren. Für diese Operation werden Sensormodelle (z.B. Kameramodell, Laser-Scanner-Modell) verwendet. Die Komprimierung wird eingesetzt, um Speicherplatz oder Kapazitäten eines Übertragungskanals einzusparen. Diese Komprimierung ist stets mit einer Codierung verknüpft. Diese ordnet z.B. den quantisierten Koeffizienten einer Karhunen-Loève Transformation (Komprimierungstransformation) den Huffman-Code zu oder es wird das digitale Bild ohne vorheriger Komprimierungstransformation direkt codiert (z.B. Lauflängen- und Richtungscodierung). Manchmal (Nachrichtentechnik) wird der Quantisierungsvorgang selbst auch als Codierung (z.B. Pulse Code Modulation, PCM) bezeichnet.

Verbesserungsoperationen arbeiten ohne A-priori-Wissen über die Störungsursachen (z.B. Übertragungsursachen, bewegtes Objekt), sondern versuchen, gestörte Bilder mit allgemein gültigen Operationen zu glätten oder zu verschärfen. Typische Vertreter dieser Operationsklasse sind Filteroperationen. Eine weitere gängige Verbesserungsoperation ist die Fouriertransformation (Abschnitt 2.3). Sie führt von dem Ortsraum in den Ortsfrequenzraum und erlaubt Analysen und Änderungen von Bildelementen einfacher durchzuführen, als sie im Ortsraum möglich sind.

Die Erwartungen (Extrahieren und Vergleichen) in dieser Schicht beziehen sich stets auf Parameter, daher werden sie von uns als *Vorhersagen* bezeichnet. Die Extraktionsoperationen werden häufig direkt mit den entsprechenden vorhersagenden Operationen verknüpft und als ein kombinierter Algorithmus implementiert. Die Vorhersagen für die Aufbereitung beziehen sich meistens auf mögliche statistische Korrelation von Pixeln (Zufallsfelder, Abschnitt 2.4), auf die vorhersagende Codierung (z.B. Differentielle PCM, DPCM), auf die Erwartungswerte von Filterantworten (z.B. Kalmanfilter [Jain 89]) oder über die Grenz Ortsfrequenzen im Bild.

Die Vergleichsoperationen dieser Schicht restaurieren und rekonstruieren Bilder. Das Resultatbild, das nach der Anwendung von Ver-

gleichsoperationen entsteht, wird als Bild'_B gekennzeichnet. Es dient der zweiten Schicht als Eingabebild, aus dem Bildkomponenten extrahiert werden. Man spricht von *Restaurierung*, wenn die Störungsquellen bekannt sind. Beispiele für solche Störungsursachen sind atmosphärische Turbulenzen (Satellitenbilder), Aberrationen (Öffnungsfehler) optischer Linsensysteme, Diffraktionen (Beugungen) in mikroskopischen Systemen, relative Bewegung zwischen Objekt und Kamera oder Rauschen. Diese A-priori-Kenntnis wird in entsprechenden *Störmodellen* explizit formuliert. So kann bei einer Bewegungsunschärfe die entsprechende Punktverschmierungsfunktion (Abschnitt 2.3) explizit angegeben werden oder auf der Annahme von weißem (unkorreliertem) Rauschen kann der Wienerfilter abgeleitet werden [Rosenfeld, Kak 82].

Rekonstruktion

Ein Bild wird rekonstruiert, wenn mit Hilfe von "Projektionen" Wirkungsquerschnitte ermittelt werden, die beispielsweise einen Schnitt durch eine Gewebeschicht in einer Tomografieaufnahme darstellen. Eine "Projektion" wird durch eine Menge von Strahlintervallen definiert, die längs eines Strahles mit Hilfe der Bildfunktion (Abschnitt 2.2) erzeugt werden. Die Rekonstruktion wird vor allem für die medizinische Diagnose (Tomografie, Ultraschall) verwendet. Als Rekonstruktionsmodelle werden die physikalischen Eigenschaften der einfallenden Strahlen (z.B. Röntgenstrahlen, Synchrotronstrahlung) und die Reflexionseigenschaften der einzelnen internen Körperbestandteile (Gewebe, Knochen) verwendet. Die Vorhersagen für die vergleichenden Operationen beschreiben entweder die wahrscheinlichsten Störquellen bzw. Pixelkorrelationen (Restaurierung) oder den häufigsten Typ der Wirkungsquerschnitte.

Bildmodelle

Für die Anwendung der genannten aufbereitenden und vergleichenden Operationen benötigt man *Bildmodelle*. Für die extrahierenden und die entsprechenden Vorhersageoperationen werden diese Bildmodelle dargestellt durch Teilbereiche der Sensorik (z.B. Verzerrungen), der Signaltheorie (Rauschen, Filter), der Codierungstheorie, der Übertragungstheorie (z.B. Kanalkapazitäten) und der Geometrie (z.B. projektive Geometrie). Für die vergleichenden Operationen benutzt man als Bildmodelle meistens Störmodelle und Rekonstruktionsmodelle.

Komponentenerkennung

Die zweite Schicht erfüllt die Aufgabe der *Komponentenerkennung* (genauer Objektkomponentenerkennung). Sie enthält alle Operationen und Modelle (abstrakte Datentypen), die zur Extraktion und Erkennung von Objektkomponenten notwendig sind. Die Darstellung, die nach der Extraktion von Bildkomponenten aus dem Bild Bild'_B entsteht, wird als *Komponentenbild* (Komp_B) bezeichnet. Die einzelnen Objektkomponenten im Bild nennen wir *Bildkomponenten*. Die Modelle dieser Schicht werden als *Komponentenmodelle* (Komp_M) bezeichnet. Die resultierende

Darstellung in dieser Schicht heißt $Komp_B$. Sie beschreibt alle erkannten Bildkomponenten.

Intern ist diese Schicht nochmals zweigeteilt. Die Teilschicht der Segmentierung liegt vor der eigentlichen Schicht der Komponenten-erkennung. Die entsprechende Teilmenge der Komponentenmodelle wird von den Segmentierungsmodellen gebildet.

Ein Bildprimitiv bildet die erste Aggregationsstufe aus klassifizierten Pixeln (z.B. Kantenpunkte). Es wird als Baustein für komplexere Bildstrukturen eingesetzt. Typische Bildprimitive sind Kanten und Bereiche. Präziser ausgedrückt ist ein *Bildprimitiv* eine Funktion, die auf klassifizierten Pixeln definiert wird und deren Wertemenge eine Teilmenge der ursprünglichen Definitionsmenge ist. Aufgabe der *Segmentierung* ist es vor allem, Bildprimitive mit Hilfe von klassifizierten Pixeln zu erzeugen. Eine Segmentierung zerlegt ein Bild in Bereiche, aus denen Bildkomponenten extrahiert werden können.

Bildprimitiv,
Segmentierung

Typische Extraktionsoperationen in der Segmentierungsschicht sind Pixelklassifikationen wie z.B. die topografische Pixelklassifikation (z.B. Grat, Kamm, Tal, Hang, Abschnitt 4.5.2) oder Kantenpunktextraktionen (z.B. Kantenoperatoren, Laplace-Gauß-Operator, Abschnitt 4.2) und Kantenverdünnungen bzw. Konturverbesserung (z.B. durch Nonmaxima-Unterdrückung oder durch Relaxation). Übliche Vergleichsoperationen der Segmentierung sind Konturpunktverkettung (z.B. Hough-Transformation oder wiederum durch Relaxation [Rosenfeld, Kak 82]), Verkleinerungen oder Vergrößerungen von Flächen (Flächenwachstum, region growing).

Segmentierungs-
operationen

Die *Segmentierungsmodelle*, die für die Extraktionsoperationen verwendet werden, beschreiben Schemata zur Pixelklassifikation, Modelle von Bildprimitive wie Kantenprofile (z.B. Sprungkanten, Abschnitt 2.4.1), Bereichsmodelle (z.B. Homogenitäten von Bereichen), fotometrische Beschreibungen (z.B. spektrale Empfindlichkeiten, Abschnitt 3.3) und Einschränkungen in der Farbdarstellung (Abschnitt 3.4). Typische Merkmale, die dafür zum Einsatz kommen, sind Kantenmerkmale (z.B. Richtung) und Texturmerkmale (z.B. Texturfrequenz). Die Modelle, die bei Vergleichsoperationen insbesondere für die Konturpunktverkettung eingesetzt werden, beziehen sich auf die Parametrisierung von Konturen (z.B. Akkumulatorfeld für die Hough-Transformation), auf die heuristische Funktion bei der A*-Suche nach Konturstücken oder auf die Kurvengüte (z.B. erlaubte Krümmung) für die dynamische Programmierung [Bässman, Besslich 89]. Diesen Segmentierungsmodellen für die Vergleichsoperationen ist gemeinsam, daß sie die Stetigkeit von Grauwertänderungen (Konturlinien) und der Kurvenkrümmungen als Merkmale verwenden.

Segmentie-
rungsmodelle

Typische zwei- oder dreidimensionale Merkmale, mit deren Hilfe man Objektkomponenten erkennt, sind: Formmerkmale (z.B. Zylinderförmigkeit, bestimmte Konturkurven), Flächenmerkmale (z.B. Flächen-

Komponenten-
merkmale

größe, Krümmung, Schattierungen), volumetrische Merkmale (z.B. Höhe, Schwerpunkt, Trägheitsmomente), geometrische Merkmale (z.B. Abstand, Oberflächenorientierung als ein Primal Sketch [Marr 82]), topologische Merkmale (z.B. umschließende Flächen), Bewegungsmerkmale (z.B. Geschwindigkeit), räumliche Anordnung von Subkomponenten innerhalb einer Komponente und Position und Orientierung der Komponente im Raum [Brady 82].

Extraktions- und Vergleichsoperationen

Die Extraktionsoperationen dieser Schicht versuchen, einzelne dieser Merkmale (z.B. Abstand durch Stereoverfahren) in dem segmentierten Bild zu finden. Sie werden dabei unterstützt durch die Erwartungsangaben darüber, welche Merkmale mit welcher Wahrscheinlichkeit wo auftreten können und welche Ansichten der entsprechenden Merkmals-träger damit verknüpft sind. Die typischen Vergleichsoperationen zur Komponentenerkennung versuchen beispielsweise, die Form (unter der Verwendung eines radiometrischen Strahlungsmodelles wie Lambert-scher Strahler, Abschnitt 3.1) aus der Schattierung, der Bewegung, der Textur usw. [Horn 86] zu bestimmen, oder es wird ein direkter Vergleich zwischen einem flächenhaften bzw. räumlichen Komponentenmodell und der extrahierten Bildkomponente durchgeführt. Dieser Vergleich kann im zweidimensionalen Fall ein reiner Schablonenvergleich (Korrelation, Abschnitt 4.2.1) sein. Bei dreidimensionalen Darstellungen, wie auch manchmal bei zweidimensionalen Darstellungen, vergleicht man korrespondierende, relationale Beschreibungen oder man implementiert den Vergleich als heuristische Suche (wissensbasierter Ansatz). Die relationale (nicht-wissensbasierte) Bestimmung von Komponentenähnlichkeiten erfolgt typischerweise auf Graphen oder auf Relationalstrukturen (Teilisomorphismus), wobei wiederum Relaxationsverfahren (z.B. consistent labeling, Abschnitt 2.4.3) oder der Ansatz des dynamischen Programmierens im Zusammenhang mit Ähnlichkeitsfunktionen eingesetzt werden können (Abschnitt 8.2).

Erwartungs- und Vergleichsoperationen

Die für die Vergleichsoperationen vorgegebenen Erwartungen werden meist als Hypothesen über die Sichtbarkeit von Komponenten festgelegt. Zu den Aufgaben der Vergleichsoperationen zählen wir ferner auch die Verarbeitung von Meßungenauigkeiten (z.B. bruchstückhafte Kreissegmente, Kameraverzerrungen), die in der Extraktionsphase vorhanden sind, um einen möglichst eindeutigen Merkmalswert (z.B. Kreisradius) etwa durch Approximationen an die Bildmerkmalswerte zu berechnen. Die Erwartungs- und Vergleichsoperationen werden häufig auch zusammengefaßt. Die Ansätze der Relaxation (Iteration) und des dynamischen Programmierens eignen sich insbesondere im Hinblick auf eine effiziente Implementierung hierfür besonders gut. In beiden Ansätzen wird der Erwartungswert mit Hilfe von Wahrscheinlichkeiten ausgedrückt, die die Zuordnung eines Bildelementes (Bildmerkmal, Bildkomponente) zu einem Modellmerkmal quantifizieren. So kann beispielsweise das Relaxationsverfahren nicht nur für die iterative Segmen-