

## Kapitel 2

# Rechnersehen

Mit *Rechnersehen* wird die Konstruktion von bedeutungsvollen Beschreibungen realer Objekte aus Bildern bezeichnet. In diesem Kapitel soll ein kurzer Überblick über das Gebiet Rechnersehen gegeben werden, ohne die einzelnen Methoden detailliert darzustellen. Eine etwas ausführlichere Übersicht bieten [CF82], sowie [CD85]. Detailliertere Informationen über die einzelnen Bildverarbeitungsmethoden bietet [BB82].

Rechnersehen untergliedert sich in mehrere Vorgänge, mit denen Bilddaten verarbeitet werden. Die Grenzen zwischen den einzelnen Bereichen, sowie deren Einteilung und Benennung sind in der Literatur sehr verschieden. Hier soll zwischen *early processing*, Bildsegmentierung, Repräsentation von zweidimensionalen Formen<sup>1</sup> und Wissensrepräsentation und Nutzung unterschieden werden.

### 2.1 Early Processing

*Early processing* bezeichnet den ersten Schritt bei der Verarbeitung eines Bildes. Ziel dieses Schrittes ist es, die in einem Rohbild vorliegende Information aufzubessern. Dazu wird eine Glättung und Filterung eines Grauwertbildes (oder Farbbildes) durchgeführt. Zur Unterdrückung von Rauschen wird das Bild zunächst geglättet, das heißt, es wird für jeden Punkt das Mittel über die Intensität der benachbarten Punkte gebildet. Um Intensitätsänderungen in einem Bild festzustellen, werden benachbarte Intensitätswerte verglichen. Ist ihre Differenz groß, so liegt eine starke Intensitätsänderung vor. Um aus einem Bild Kanten extrahieren zu können, wird dieses zunächst geglättet und dann die Kantenübergänge verstärkt. Im Folgenden werden einige Verfahren des *early processing* oder *preprocessing* angeführt.

---

<sup>1</sup>Stereosehen und die Repräsentation von dreidimensionalen Formen wird hier nicht erläutert.

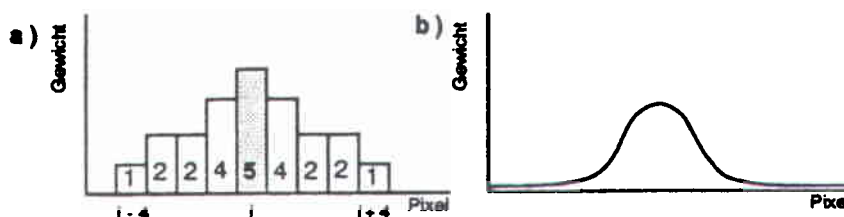


Abbildung 2.1: Gewichtung der umgebenden Pixel zur Glättung

### 2.1.1 Gauß - Filter

Ein Bild enthält häufig Rauschen, d.h. zufällige Störungen einzelner Intensitätswerte, welches für die weitere Verarbeitung des Bildes störend sein kann. Aus diesem Grund wird als erstes versucht dieses Rauschen zu unterdrücken. Durch Ersetzen des Intensitätswertes eines Pixels durch den Mittelwert der Intensitäten der Nachbarpixel, können zufällig auftretende Intensitätsänderungen abgeschwächt werden.

Die bei der Berechnung des neuen Intensitätswertes in Betracht gezogenen Pixel können entsprechend ihrer Entfernung zu dem Bezugspixel gewichtet werden. Dabei werden näher liegende Pixel mit einem höheren Gewicht versehen. Abbildung 2.1a zeigt eine mögliche Gewichtung der umgebenden Pixel. Dabei wurde eine Pixelzeile des Bildes betrachtet. Wenn mehr Punkte zur Verfügung stünden, könnte eine kontinuierliche Kurve in Form einer Gauß-Kurve (Abbildung 2.1b) zugrundegelegt werden anstatt diskrete Werte zu betrachten.

Da bei der Glättung das zweidimensionale Pixelarray betrachtet wird, werden nicht nur die benachbarten Pixel einer Zeile, sondern auch die in der Spalte benachbarten Pixel in die Mittelwertberechnung einbezogen.

### 2.1.2 Laplace-Ableitung

Um verschwommene Konturen zu verbessern, müssen die Kanten verstärkt werden. Dies kann durch Subtraktion der zweiten Ableitung von der Intensitätsfunktion erreicht werden. Die erste Ableitung gibt an, wie stark sich die Intensität längs einer Linie über das Grauwertbild ändert, während die zweite Ableitung die Stärke der Intensitätsänderung angibt. Abbildung 2.2 verdeutlicht dies an einem Beispiel. Die Kurve gibt den Verlauf der Helligkeit entlang einer Linie senkrecht zur Kante wieder. Der Laplace Operator bildet die Ableitung des zweidimensionalen Grautonfeldes. Zur Anwendung muß der Laplace Operator diskretisiert werden. Dazu bildet man

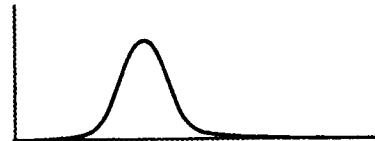
a) Kantenübergang in einem  
Grauwertbild



b) Helligkeitsverlauf



c) 1. Ableitung



d) 2. Ableitung

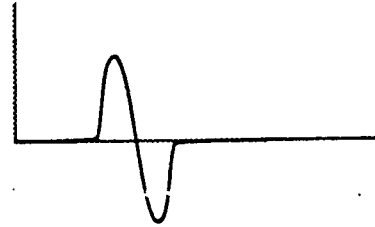


Abbildung 2.2: Differenzierung einer Intensitätsfunktion

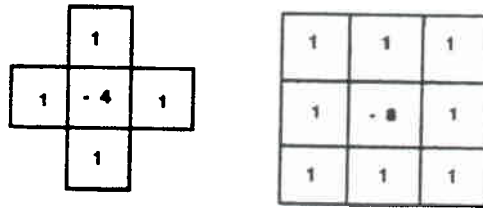


Abbildung 2.3: Laplace Filter

zunächst die vereinfachten Ableitungen:

$$\begin{aligned} f_x(i, j) &= f(i, j) - f(i - 1, j) \\ f_y(i, j) &= f(i, j) - f(i, j - 1) \end{aligned}$$

Dabei stellt  $f$  die Intensität eines Bildpunktes dar, dessen Koordinaten durch  $i$  und  $j$  gegeben sind. Die zweiten Ableitungen werden auf die gleiche Weise gewonnen.

$$\begin{aligned} f_{xx}(i, j) &= f_x(i + 1, j) - f_x(i, j) \\ &= f(i + 1, j) + f(i - 1, j) - 2f(i, j) \\ f_{yy}(i, j) &= f_y(i, j + 1) - f_y(i, j) \\ &= f(i, j + 1) + f(i, j - 1) - 2f(i, j) \end{aligned}$$

Daraus ergibt sich die diskrete Version des Laplace Operators:

$$\begin{aligned} \nabla^2 f(i, j) &= f_{xx}(i, j) + f_{yy}(i, j) \\ &= f(i, j + 1) + f(i, j - 1) + f(i + 1, j) + f(i - 1, j) - 4 \cdot f(i, j) \end{aligned}$$

Abbildung 2.3 zeigt den Laplace Operator, sowie einen weiteren Operator, der zusätzlich noch die diagonal benachbarten Pixel betrachtet.

### 2.1.3 Der mexikanische Hut

Für die Durchführung der beiden Schritte Differenzieren und Glätten wurden bereits mehrere Verfahren entwickelt. Eines der einfachsten Verfahren ist durch die Zusammenfassung der beiden Schritte gegeben. Durch die Laplace-Ableitung einer Gaußfunktion erhält man einen Filter, mit dem man das Bild falten<sup>2</sup> kann. Die Laplace-Ableitung verwandelt die Gaußkurve in eine Kurve, die einem mexikanischen Hut ähnelt (siehe auch Abbildung 2.4). Da nicht nur eine Pixelreihe, sondern ein Pixelfeld betrachtet wird, muß man sich die Kurve noch um ihre senkrechte Mittelachse gedreht vorstellen.

<sup>2</sup>Falten bedeutet, daß jedes Pixel durch das gewichtete Mittel seiner umgebenden Pixel ersetzt wird.

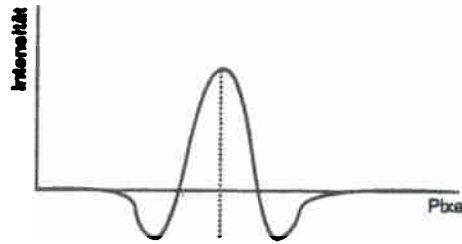


Abbildung 2.4: Ein mexikanischer Hut als Filter

Am Rand der Kurve entsteht eine *Krempe*. Helle Punkte, die in diesem Bereich um das Pixel liegen, bringen negative Werte in die Berechnung des Mittelwertes für das betrachtete Pixel ein. Das Resultat der Faltung ist ein Feld aus positiven und negativen Zahlen, wobei die Nullstellen des Feldes denjenigen Stellen im Bild entsprechen, an denen sich die Intensität am stärksten ändert.

## 2.2 Bildsegmentierung

Nach der Aufbereitung des Bildes, gilt es nun, das Bild in Teilbereiche zu untergliedern, die zu Objekten, bzw. Teilbereichen von solchen gehören. Dies kann durch Betrachtung der Farbe, der Helligkeit oder der Textur<sup>3</sup> geschehen. Bei der Vorgehensweise kann zwischen kantenorientierten und bereichs- oder regionenorientierten Segmentierungsstrategien unterschieden werden.

### 2.2.1 Kantenextraktion

Kantenorientierte Verfahren beruhen auf der Anwendung eines lokalen Kantenoperators auf das Bild und einem nachfolgenden Verarbeitungsschritt zur Detektion signifikanter Kanten. Der oben beschriebene Laplace-Operator stellt einen solchen lokalen Kantenoperator dar. Das Ergebnis der Operation ist ein Bild, bei dem die Nulldurchgänge<sup>4</sup> eine hohe Wahrscheinlichkeit für eine Kante angeben. Durch eine Schwellwertoperation können somit Objektgrenzen bestimmt werden. Dies ist

<sup>3</sup>Eine Textur ist eine Anordnung von kleinen (feinkörnigen) Mustern, die in einer bestimmten Struktur angeordnet sind. Regionen mit gleicher Textur erscheinen homogen. Bilder von Stoffen weisen beispielsweise eine Textur auf.

<sup>4</sup>Bei anderen Operatoren als dem Laplace-Operator können auch die Maxima ausschlaggebend sein. (bsp. Sobel-Operator)

jedoch bereits bei mäßig komplexen Bildern schwer, da die Konturen meist nicht geschlossen sind.

### 2.2.2 Regionenanalyse

Regionenorientierte Verfahren sind nicht an ein starres, lokales Fenster gebunden, welches über das Bild bewegt wird, wie das bei der Kantenextraktion der Fall ist. Sie operieren auf beliebig geformten zusammenhängenden Pixelmengen, den Regionen. Eine mögliche Operation regionenorientierter Verfahren sind die Verschmelzung zweier Regionen (*merge*) falls ein Ähnlichkeitskriterium<sup>5</sup> erfüllt ist. Außerdem kann eine Region aufgespalten werden (*split*), falls ein Homogenitätskriterium nicht mehr erfüllt wird. Die kleinste erlaubte Region ist ein Pixel.

Die Strategien der regionenorientierten Bildanalyse werden in zwei Klassen unterteilt. Wird eine einzige Region, also das komplette Bild, als Ausgangspunkt genommen, so wird die Methode als *top down* Verfahren bezeichnet. Diese Verfahren beruhen auf einem Homogenitätstest, durch den Regionen immer weiter in kleinere Regionen zerlegt werden. Eine zweite Klasse bilden die *Wachstumsverfahren*. Aufgrund eines Ähnlichkeitskriteriums wird entschieden, ob Regionen zusammengefaßt werden dürfen. Durch dieses Verfahren können Regionengrenzen sehr gut erfaßt werden. Außerdem existieren sogenannte *split-and-merge* Verfahren, die beide oben beschriebenen Verfahren vereinen. Nähere Informationen zur regionenorientierten Bildbeschreibung können [Fra89] entnommen werden.

## 2.3 Repräsentation von zweidimensionalen Formen

Eine der wichtigsten Informationen, die aus einem Bild gewonnen werden müssen, ist die Kontur der Objekte, die im Bild enthalten sind. Durch die zwei zuerst angeführten Verfahren *early processing* und *Bildsegmentierung* werden die Pixel bestimmt, die zu einer Kante gehören können. Es wurde jedoch noch keine Form der Kante, sondern lediglich Punktmengen bestimmt, die eine Kante bilden. Kanten haben verschiedene Erscheinungsformen. Für jede dieser Erscheinungsformen muß eine Repräsentation gefunden werden. Weitere Informationen zu diesem Bereich sind auch in [Bar87] nachzuschlagen.

---

<sup>5</sup> Dieses Ähnlichkeitsmaß ist in Abhängigkeit von der betrachteten Sensorinformation zu definieren. Es kann sich dabei um ähnliche Textur zweier Regionen, einen ähnlichen Helligkeitswert oder ähnliche Farbwerte handeln.

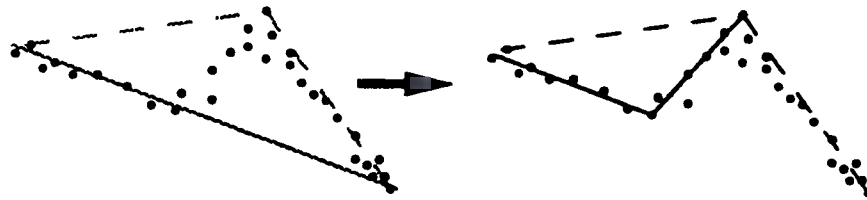


Abbildung 2.5: Iterative Bestimmung der Eckpunkte

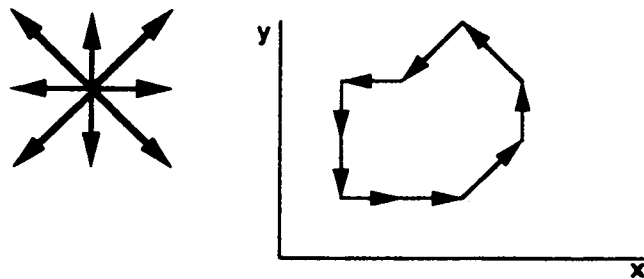


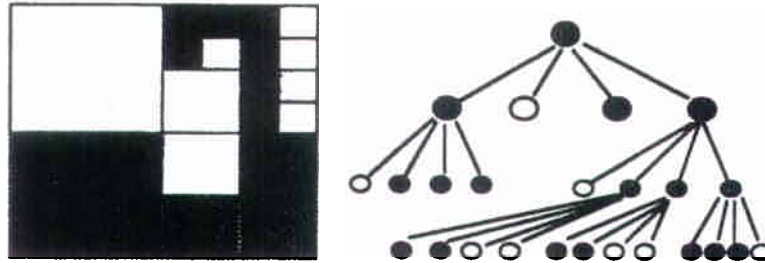
Abbildung 2.6: Beispiel einer chain code-Beschreibung

### 2.3.1 Linienzüge als Punktliste

Zu Bestimmung von Eckpunkten einer Kante, sowie zur Detektion von Kreisen und Krümmungen wurden mehrere Methoden entwickelt. Abbildung 2.5 zeigt ein Beispiel einer einfachen Methode zur Bestimmung der Eckpunkte. Zunächst werden die beiden Punkte der Kante verbunden, die am weitesten entfernt liegen. Der Punkt mit dem größten Abstand zu der entstehenden Linie wird in die Eckpunktliste aufgenommen und das Verfahren wird auf die beiden neuen Linien angewandt, bis ein Schwellwert unterschritten wird.

### 2.3.2 Chain codes

*Chain codes* bestehen aus Liniensegmenten, die auf einem fixen Raster liegen und eine begrenzte Anzahl an Orientierungen haben können. Nur ein Anfangspunkt wird durch seine Koordinaten im Bild festgelegt, die anderen Punkte werden allein durch eine Kette von Richtungsangaben bestimmt. Abbildung 2.6 zeigt ein Beispiel dazu.

Abbildung 2.7: Die Repräsentationsform *quad tree*

### 2.3.3 B-Splines

Kurven können durch eine Interpolation angegeben werden. Dabei werden lediglich die Stützpunkte gespeichert, aus denen die Kurve gewonnen werden kann. Ein B-Spline wird durch ein Polygon repräsentiert, welches die Kurve des B-Splines festlegt (*guiding polygon*). Splines bilden eine günstige Darstellungsform, da sie die Möglichkeit bieten, komplexe Formen zu modellieren.

### 2.3.4 Regionenrepräsentation

Eine Möglichkeit zur Repräsentation von Regionen bieten *quad trees*. Abbildung 2.7 zeigt den Aufbau dieser Repräsentationsform am Beispiel. Das Bild wird zunächst in vier gleiche Teile aufgeteilt. Jedem dieser Teile kann ein Wert schwarz, weiß oder grau zugeordnet werden. Sind alle Pixel eines Teils schwarz, so ist der gesamte Bereich schwarz. Gleiches gilt für weiß. Sind in einem Bereich schwarze sowie weiße Komponenten, so wird der Bereich als grau bezeichnet und auf die gleiche Weise unterteilt. Es ergibt sich ein Baum, dessen Blätter entweder weiß oder schwarz sind und dessen Wurzel das Bild repräsentiert.

### 2.3.5 Eigenschaften von Konturen

Eigenschaften, die aus der Kontur eines Objekts gewonnen werden können, sind beispielsweise die Fläche, der Schwerpunkt, die Kompaktheit. Die Fläche eines Objekts kann aus der Konturbeschreibung des Objekts gewonnen werden. Die Kompaktheit wird durch das Verhältnis Umfang zu Fläche bestimmt.



## 2.4 Repräsentation und Nutzung von Wissen

### 2.4.1 Repräsentationsform

Um eine durch ein Bild dargestellte Szene zu verstehen, muß ein Modell von der Welt existieren in der diese Szene auftreten kann. Dazu wird eine Wissensbasis angelegt, in der Wissen über die Struktur der Welt explizit dargestellt wird. Eine häufig verwendete Darstellungsform im Bereich Rechnersehen stellt das *Semantische Netz* dar. Semantische Netze sind gut geeignet um komplexe Informationen zu strukturieren. Weitere Repräsentationsformen sind Grammatiken und Regelsysteme.

### 2.4.2 Matching

Ist ein Modell der Welt vorhanden, so müssen die modellierten Objekte in einem Bild gefunden werden. Dazu wird aus der Bildinformation eine Repräsentation abgeleitet, die mit der Repräsentationsform des Modells verglichen werden kann, so daß Gemeinsamkeiten und partielle Ähnlichkeiten zwischen Modell und Bild gefunden werden können.

### 2.4.3 Inferenz

Aus gefundenen Teilstrukturen kann mit Hilfe von Inferenzmechanismen (Resolution, Produktionssysteme,...) abgeleitet werden, welche Teilstrukturen an welchem Ort in der Szene zu finden sind. Dadurch kann eine aufwändige Durchsuchung des gesamten Bildes nach bestimmten Objekten vermieden werden. Aus einer Anzahl von Teilinformationen über Objekte kann auf weitere Eigenschaften geschlossen werden. Die Zuordnung zwischen einem Bild und seiner Bedeutung geschieht ebenfalls über den Inferenzmechanismus.