Markoff-Felder als Hilfsmittel der Bildverarbeitung

Diplomarbeit
von
Ulrik Brandes
Matr.-Nr. 171 830

Lehrstuhl für Angewandte Mathematik,
insbesondere Informatik
RWTH Aachen
1994

Referent und Betreuer: Prof. Dr. Walter Oberschelp
Koreferent: Prof. Dr. Hermann Ney

Meinen Eltern gewidmet

Vorwort

Die Welt ist meine Vorstellung.
A. SCHOPENHAUER

Bilder haben die Menschheit seit jeher beschäftigt. Möglicherweise *bilden* sie sogar die Grundlage unseres Seins, da sie nicht nur Mittler zwischen Umwelt und Existenz (Sinn-Bilder), sondern auch elementarste Hilfsmittel des Denkens zu sein scheinen.

Eine Faszination ganz anderer Art geht von der versuchten Beherrschung des Zufalls aus, die das menschliche Streben nach Anhaltspunkten im Unsicheren offenbart. Promethische Wissbegierde und Furcht vor dem Unbekannten sind nur zwei der sich anbietenden Auslegungen desselben.

Neben der Betreuung bin ich Prof. Oberschelp daher besonders für das mir gestellte Thema dankbar, in dem die schon für sich genommen reizvollen Gebiete zusammenfließen und bei dessen Bearbeitung er mir alle Freiheiten ließ. Prof. Ney danke ich für die bereitwillige Übernahme des Koreferates.

An der Entstehung dieser Arbeit waren zahlreiche weitere Personen mittel- oder unmittelbar beteiligt. Teile des Manuskriptes haben durchgesehen Dr. Volker Penner, Dipl.-Inform. Bernadette Lipinski, Dipl.-Math. Jürgen Mattfeldt, Dipl.-Inform. Frank Weiler und cand. Phys. Irene Merk. Zusätzlich zum vorgenannten Personenkreis erteilten mir Prof. Rudolf Mathar, Dipl.-Ing. André Kaup und Dipl.-Phys. Thorsten Brikey bereitwillig Auskunft auf diverse Fragen. Prof. Stuart Geman wies mich auf die Arbeit von Robert Azencott hin.

Unterstützung durch Literatur und Bildmaterial erhielt ich von Dipl.-Ing. Thomas Lehmann, Dipl.-Ing. Erich Pelikan und wiederum F. Weiler, B. Lipinski, A. Kaup und I. Merk. Cand. Inform. Roland Schindler machte mich mit der SIMPLEX-Bibliothek bekannt, und cand. phil. Ursula Höppe hat die Arbeit Korrektur gelesen. Ihnen allen gilt mein herzlichster Dank.

Hiermit versichere ich, daß ich die vorliegende Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, den 11. Juli 1994 Ulrik Brandes

Inhaltsverzeichnis

Kapitel 1

Einleitung

Das letzte, was man findet,
wenn man ein Werk schreibt,
ist, womit man beginnen soll.
B. Pascal

Gegenstand dieser Arbeit ist ein stochastisches Modell für digitalisierte und quantisierte (endliche) Bilddaten. Es soll gezeigt werden, daß die Bildbeschreibung durch Markoff-Felder ein tragfähiges Konzept ist, das auf plausible Modellierungen und geschlossene Algorithmen zur Verarbeitung von Bildern führt.

Bilder sind im Rahmen dieser Ausarbeitung Wiedergaben von Umweltsituationen, wodurch der Bildbegriff jedoch nicht notwendig auf visuelle Eindrücke beschränkt wird. Ungeachtet der Beispiele und Begriffsbildungen sollte immer berücksichtigt werden, daß es sich ebensogut um akustische, thermische oder sonstige Meßgrößen handeln kann. Selbst nicht-physikalische Umweltinformation kann weiterhin als Bild aufgefaßt werden, sofern sie sich als räumlich angeordnete Menge von Zuständen repräsentieren läßt.

Bild*verarbeitung* umfaßt prinzipiell alle Arbeitsschritte von der Erzeugung eines Bildes über die Veränderung der dabei angefallenen Daten bis hin zur Anzeige des Ergebnisses oder gar zum Fällen einer Entscheidung aufgrund von Bildinhalten.

In der folgenden Definition werden die bezüglich der zugrundeliegenden Daten getroffenen Vereinbarungen konkretisiert und die grundlegenden Bezeichnungen festgelegt:

Definition 1.1 (Grundbegriffe)
Es seien $N, M, G \in \mathbb{N}$ natürliche Zahlen.

- *Es seien* $I = \{1, \ldots, N\}$ *die Menge der* Spalten- *und* $J = \{1, \ldots, M\}$ *die Menge der* Zeilenindizes. $L = I \times J$ *heißt* Gitter, $(i, j) \in L$ *heißt* Bildpunkt, Gitterpunkt *oder* Pixel.

- $\Gamma = \{0, \ldots, G-1\}$ *heißt* Grauwertmenge *und* $g \in \Gamma$ Grauwert.

- $\Gamma^L := \{(x_{1,1}, \ldots, x_{N,M}) | x_{ij} \in \Gamma, \, (i, j) \in L\}$ *heißt* Bildraum, *und jedes* $x \in \Gamma^L$ *ist ein* Bild.

Für die Elemente des Bildraumes kommt es auf die Reihenfolge der Indizes aus L nicht an, da eine Anordnung der Bildpunkte nicht erforderlich ist. Im Prinzip entspricht Definition 1.1 natürlich der für digitalisierte Bilder üblichen, in der an die Stelle des Vektors $x \in \Gamma^L$ eine *Ortsbereichsfunktion* $f : L \to \Gamma$ tritt [OD91].

Typischerweise ist die Anzahl $|L|$ der Bildpunkte groß, z.B. $N = M = 1024$, und ein Bild $x \in \Gamma^L$ schwierig zu interpretieren. Für eine kompakte Beschreibung charakteristischer Bildeigenschaften wurden daher Kenngrößen definiert, welche die Datenfülle auf wenige Parameter reduzieren und/oder in eine Form bringen, die das Ablesen von Merkmalen erleichtert. In den folgenden beiden Abschnitten sind einige Beispiele zusammengestellt. Entsprechend definierte Größen werden in [Jäh91] und [Hab87] ausführlicher besprochen.

1.1 Statistik 1. Ordnung

Generelle Aussagen über Kontrastumfang, Bildhelligkeit oder vorherrschende Grauwerte liefert eine einfache Zählung der Grauwertvorkommen. Man erhält eine Zuordnung von Anzahlen zu Grauwerten, das sogenannte *Histogramm* $h : \Gamma \to \mathbb{N}_0$ mit

$$h(g) := |\{\, (i, j) \in L \,|\, x_{ij} = g \,\}|.$$

Offensichtlich gilt $\sum_{g \in \Gamma} h(g) = N \cdot M$, so daß bei einer Gleichgewichtung aller Pixel eine diskrete Wahrscheinlichkeitsverteilung auf den Grauwerten definiert ist durch die Zähldichte $p = (p_0, \ldots, p_{G-1})$ mit

$$p_g := \frac{h(g)}{N \cdot M} \qquad g \in \Gamma.$$

Die aus der Stochastik bekannten Kenngrößen von Verteilungen übertragen sich nun auf die Grauwertverteilung. Zur Veranschaulichung werden drei dieser Größen zusammen mit einer groben Interpretation angegeben:

$$\begin{array}{rcllll} \mu & := & \sum_{g\in\Gamma} g\cdot p_g & \text{Mittelwert} & \text{(Helligkeitseindruck)} \\ \sigma^2 & := & \sum_{g\in\Gamma} (g-\mu)^2 p_g & \text{Varianz} & \text{(Kontrastumfang)} \\ H(p) & := & \sum_{g\in\Gamma} p_g \cdot \log \frac{1}{p_g} & \text{Entropie} & \text{(Beschreibungsaufwand}^1\text{)} \end{array}$$

Hat man eine Serie von Bildern (Ensemble) gegeben, so kann für jeden einzelnen Bildpunkt eine eigene Zähldichte $p(i,j) = (p(i,j)_0, \ldots, p(i,j)_{G-1})$ auf der Grauwertmenge definiert werden, indem bezüglich des Ensembles die relativen Häufigkeiten der Grauwerte jeweils nur im Punkt (i,j) berechnet werden. Ersetzt man p durch $p(i,j)$, führt dies auf punktabhängige Kenngrößen $\mu(i,j)$ etc.

Obige (und weitere) Größen sind besonders hilfreich in der Bildverbesserung. Durch Manipulation des Histogramms mittels *Punktoperationen* (Abbildungen von Γ nach Γ) kann der Eindruck der Bildqualität häufig erhöht werden, indem z.B. die auftretenden Grauwerte besser über die ganze Skala verteilt werden (*Histogrammspreizung*).

Weil jedes Pixel isoliert betrachtet wird, enthalten Histogramme und die daraus abgeleiteten Kenngrößen allerdings keinerlei Information über die Bildstruktur. Der nächste Schritt besteht deshalb in der Verknüpfung der Grauwerte an verschiedenen Bildpunkten.

1.2 Statistik 2. Ordnung

Ein erster Ansatz, Aussagen über die Struktur eines Bildes zu machen, besteht in der Auswertung von Abhängigkeiten zwischen Grauwerten an je zwei Pixeln. Für einen Verschiebungsvektor $(d_i, d_j) \in \mathbb{Z}^2$ benötigt man als erstes ein Wahrscheinlichkeitsmaß für das gemeinsame Auftreten zweier Grauwerte an zwei um den Vektor (d_i, d_j) auseinanderliegenden Bildpunkten.

Man definiert daher als empirische Größe zunächst die Häufigkeit des Auftretens eines Grauwertpaares (g,h) an den Pixeln (i,j) und $(i+d_i, j+d_j)$,

$$h_{d_i d_j}(g,h) := |\{(i,j) \in L \,|\, (i+d_i, j+d_j) \in L,\ x_{ij} = g,\ x_{i+d_i,j+d_j} = h\}|,$$

und verwendet dann als Zähldichte die relativen Häufigkeiten

$$p^2(d_i,d_j)_{gh} := \frac{h_{d_i d_j}(g,h)}{|\{(i,j) \in L | (i+d_i, j+d_j) \in L\}|}.$$

[1] Als Beschreibungsaufwand kann die durchschnittlich zur Codierung eines Bildpunktes benötigte Bitlänge angesehen werden [Hab87]. Dies gilt jedoch nur für den Fall, daß keine strukturelle Zusatzinformation vorliegt, da sich beispielsweise ein Schachbrett (trotz maximaler Entropie) sehr einfach codieren läßt.

Die Matrizen $C_{d_i d_j} = \left(h_{d_i d_j}(g,h)\right)_{g,h\in\Gamma}$ werden als *co-occurrence*-Matrizen bezeichnet. Man kann sie für unterschiedliche Verschiebungsvektoren kombinieren und aus ihnen wiederum Kenngrößen ableiten (für Beispiele siehe [Har79], für eine Anwendung z.B. [Wei93]).

Eine abgeleitete Kenngröße, welche die Grauwerte an den Punkten (i,j) und $(i+d_i, j+d_j)$ verknüpft,ist die *Autokorrelation*

$$r_{d_i,d_j} := \sum_{g,h\in\Gamma} g \cdot h \cdot p^2(d_i,d_j)_{gh}.$$

Besser abschätzen läßt sich der Grad der gegenseitigen Beeinflussung, wenn man die Grauwerte vor der Verknüpfung zentriert, also jeweils den Mittelwert abzieht. Daraus ergibt sich die *Autokovarianz*

$$\sigma_{d_i d_j} := \sum_{g,h\in\Gamma} (g-\mu)\cdot(h-\mu)\cdot p^2(d_i,d_j)_{gh},$$

die für $(d_i,d_j) = (0,0)$ natürlich gerade der Varianz entspricht.

Analog zu oben lassen sich für Bildserien auch die gemeinsamen relativen Häufigkeiten punktabhängig bestimmen. Wird $p^2(d_i,d_j)$ durch $p^2(i,j;d_i,d_j)$ ersetzt, so ergibt sich für die Autokorrelation im Punkt (i,j) eines Bildes x des Ensembles

$$r(i,j)^{xx}_{d_i d_j} := \sum_{g,h\in\Gamma} g \cdot h \cdot p^2(i,j;d_i,d_j)_{gh}$$

und enstprechend die Autokovarianz $\sigma(i,j)^{xx}_{d_i d_j}$.

Man spricht von *Auto*korrelation und -varianz, weil es sich um Größen handelt, welche die Grauwerte eines gegebenen Bildes *untereinander* in Beziehung setzen. Für zwei verschiedene Bilder x und y des Ensembles erhält man die *Kreuzkorrelation* im Punkt (i,j)

$$r(i,j)^{xy}_{d_i d_j} := \sum_{g,h\in\Gamma} g \cdot h \cdot p^2(i,j;d_i,d_j)_{gh}$$

und entsprechend die *Kreuzkovarianz* $\sigma(i,j)^{xy}_{d_i d_j}$.

Neben den aufgeführten werden je nach Anwendungsfall noch weitere Größen definiert, die Aussagen über interessierende Bildstrukturen gestatten. Dennoch sind alle diese Größen bestenfalls Maßzahlen für den Grad des Vorhandenseins bestimmter Bildeigenschaften. Hier sollen darüber hinausgehend Bilder als Ganzes mit stochastischen Mitteln beschrieben werden.

1.3 Inhaltsübersicht

Die oben vorgestellten Kenngrößen für Bilder sind sicherlich stochastischer Natur. Es wäre dennoch überzogen, schon von stochastischer Bildverarbeitung zu sprechen, weil die letztendliche Grundlage der auf diesen Größen basierenden Algorithmen meist doch wieder das deterministische Bildmodell der Ortsbereichsfunktion ist. Da die Leistungsfähigkeit von Bildverarbeitungsalgorithmen sehr stark von der Klasse der zu behandelnden Bilder abhängt, ist aber gerade das diese Klasse charakterisierende Modell von entscheidender Bedeutung [ChJ93, Kapitel 1].

Eine der wesentlichen Eigenschaften der Modellierung durch Markoff-Felder ist, daß die daraus hergeleiteten Berechnungsziele Optimalitätskriterien genügen, bei denen in der Herleitung der Algorithmen erkennbar wird, an welchen Stellen Praktikabilitätskompromisse eingegangen werden.

Weil die stochastische Bildverarbeitung mit Hilfe des Markoff-Modells erst in den 80er Jahren (dann allerdings rapide) Verbreitung erfahren hat, ist die Literatur noch uneinheitlich in Bezeichnungen und Konventionen. Es wurde deshalb der Versuch unternommen, die Darstellung durch einen konsistenten Formalismus zu vereinheitlichen.

Die in bezug auf stochastische Hilfsmittel verwendeten Schreibweisen folgen weitestgehend denen aus [MP90], wo auch die hier benötigten wahrscheinlichkeitstheoretischen Grundlagen nachgelesen werden können.

Die nachfolgenden Kapitel sind wie folgt gegliedert: Im zweiten Kapitel wird das grundlegende Bildmodell über seine lokale und globale Charakterisierung eingeführt. Weiterhin wird die Äquivalenz dieser Charakterisierungen bewiesen und eine neue Teilklasse von Markoff-Feldern vorgeschlagen, die einen Kompromiß aus Allgemeinheit, praktischem Nutzen und Anschaulichkeit herstellen soll.

Das dritte Kapitel enthält Beispiele für Anwendungen des Markoff-Modells, die nach zwei Kriterien ausgewählt wurden: Zum einen sollten die zentralen Anwendungsbereiche abgedeckt, zum anderen aber auch die wichtigsten Techniken im Umgang mit dem Modell vorgestellt werden. Zu diesem Zweck folgen die Ausführungen in den Abschnitten 3.2, 3.3 und 3.4 jeweils einem beispielhaften Ansatz aus der mittlerweile breit gefächerten Literatur.

Um eine von spezifischen Problemlösungen losgelöste Untersuchung von Berechungsfragen zu ermöglichen, wurde in Kapitel 4 ein allgemeines Maschinenmodell für die wichtigste im Zusammenghang mit Markoff-Feldern verwendete Methode, das Simulated Annealing mit Zufallsfeldern, entworfen.

Auf einige Schlußbemerkungen im fünften Kapitel folgt in Anhang A noch eine kurze Beschreibung der Implementation des Berechnungsschemas aus

Kapitel 4, die zur Erstellung einiger in Abschnitt 3.1 gezeigter Texturbeispiele benutzt wurde.

Kapitel 2

Bildmodellierung durch Zufallsfelder

> **Wenn Herr K. einen Menschen liebte**
> "Was tun Sie", wurde Herr K. gefragt, "wenn Sie einen Menschen lieben?" "Ich mache einen Entwurf von ihm", sagte Herr K., "und sorge, daß er ihm ähnlich wird". "Wer? Der Entwurf?" "Nein", sagte Herr K., "der Mensch".
>
> B. Brecht

Dieses Kapitel vermittelt die Grundlagen der Verwendung von Zufallsfeldern in der Bildverarbeitung. Ausgehend von der anschaulichen Interpretation der Abhängigkeit eines Pixelgrauwertes von denen der nahegelegenen Pixel werden Markoff-Felder als stochastisches Bildmodell eingeführt. Nachfolgend wird gezeigt, daß es neben der sich in der Definition der Markoff-Felder widerspiegelnden lokalen Charakterisierung von Bildeigenschaften auch eine äquivalente globale gibt.

Schließlich wird eine Teilklasse von Zufallsfeldern angegeben, deren Elemente durch wenige, interpretierbare Parameter bestimmt sind. Daß diese Klasse andererseits groß genug ist, um von Nutzen zu sein, wird durch Einordnung der Modelle zahlreicher anderer Autoren in diese Klasse belegt.

2.1 Motivation

Bei zahlreichen Fragestellungen der Bildverarbeitung stößt man auf Einflüsse, die sich der deterministischen Beschreibung entziehen. Dies kann einerseits an der stochastischen Natur dieser Einflüsse (z.B. Rauschen), andererseits aber auch an unvollständigem Wissen über die relevanten Größen (z.B. Bildrekonstruktion aus Projektionen) liegen. Das hier verfolgte Ziel besteht darin, solche Phänomene adäquat zu modellieren und die praktische Nützlichkeit und theoretische Geschlossenheit des Modells nachzuweisen.

2.1.1 Zufallsfelder

Ein erster Ansatz, nicht nur einige Kenngrößen, sondern Bilder als Ganzes den Mitteln der Stochastik zugänglich zu machen, muß notwendigerweise darin bestehen, die feste Zuweisung von Grauwerten zu Bildpunkten mittels der Ortsbereichsfunktion durch eine wahrscheinlichkeitsabhängige Zuordnung zu ersetzen. Folgerichtig tritt an die Stelle der *Funktion* $f : L \to \Gamma$ ein *Zufallsvektor* $X = (X_{1,1}, \ldots, X_{N,M})$, dessen Komponenten den Gitterpunkten zugeordnet sind und in die Grauwertmenge abbilden. Auf diese Weise wird beschrieben, daß ein Pixel einen bestimmten Grauwert nur mit einer ihm zugeordneten Wahrscheinlichkeit annimmt.

Definition 2.1 (Zufallsfeld)
Ein Zufallsvektor $X = (X_{1,1}, \ldots, X_{N,M})$ *auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, P)$*, dessen Komponenten in die Grauwertmenge* Γ *abbilden, heißt* Zufallsfeld (RF)[1] *über dem Gitter* L*.*

Die Bildmodellierung durch ein Zufallsfeld interpretiert man so, daß jedem Bildpunkt $(i,j) \in L$ eine Zufallsvariable $X_{ij} : (\Omega, \mathcal{A}) \to (\Gamma, \mathcal{P}(\Gamma))$ zugeordnet wird, so daß (i,j) den Grauwert $g \in \Gamma$ mit Wahrscheinlichkeit $P(X_{ij} = g)$ annimmt. Man beachte, daß mit den unterliegenden Gittern und den Grauwertmengen auch die darüber definierten Zufallsfelder endlich sind.

Beispiel 2.2 (Extrembeispiele)

(i) *Es sei eine Ortsbereichsfunktion* $f : L \to \Gamma$ *gegeben. Sind die Randverteilungen* $P^{X_{ij}}$ *eines Zufallsfeldes* X *über dem Gitter* L *stochastisch unabhängig und verteilt gemäß*

$$P(X_{ij} = x_{ij}) = \begin{cases} 1 & \textit{falls } f(i,j) = x_{ij} \\ 0 & \textit{sonst,} \end{cases}$$

so ist P^X *eine Einpunktverteilung mit dem durch die Funktion* f *beschriebenen Bild* $x = (f(1,1), \ldots, f(N,M))$ *als einzigem Trägerelement.*

[1] engl.: *random field*

(ii) *Sind alle Randverteilungen $P^{X_{ij}}$ eines Zufallsfeldes stochastisch unabhängig und gleichverteilt auf Γ, so ist jedes Bild $x \in \Gamma^L$ gleichwahrscheinlich.*

Die Beispiele 2.2 sind sicher pathologisch, denn während für *(i)* kein stochastisches Modell nötig wäre, liefert *(ii)* keine brauchbare Spezifizierung der betrachteten Bildklasse. Das folgende Beispiel aber behandelt eine erste Klasse von relevanten Zufallsfeldern, die deterministisch nicht zureichend beschreibbar sind.

Beispiel 2.3 (Weißes Rauschen)
Ist ein Bild durch ein Zufallsfeld X modelliert, so läßt sich auf einfache Weise eine Beschreibung für das durch Rauschen gestörte Bild $\tilde{X}$ gewinnen, wenn das Rauschen durch seine stochastischen Eigenschaften charakterisiert ist. Handelt es sich um normalverteiltes weißes Rauschen, so kann es durch ein weiteres Zufallsfeld R beschrieben werden, dessen Komponenten stochastisch unabhängig und gemäß $\mathcal{N}(\mu, \sigma^2)$ verteilt sind. Man erhält dann aus

$$\tilde{X} = X + R \quad \textit{bzw.} \quad \tilde{X} = X \cdot R$$

eine Verteilung für das durch additives bzw. multiplikatives Rauschen gestörte Bild $\tilde{X}$. Die Verknüpfungen $+$ und $\cdot$ sind dabei komponentenweise zu verstehen.

2.1.2 Lokale Abhängigkeiten

Da natürlich auch ganz andere Bildklassen modelliert werden sollen, stellt sich die Frage, wie man zu sinnvollen Verteilungen für Zufallsfelder gelangen kann, in denen die Komponenten nicht stochastisch unabhängig sind. Dazu soll zunächst ein suggestives Beispiel gegeben werden. Abbildung 2.1 zeigt einige einfache Bilder, deren mittleres Pixel entfernt worden ist. Das fehlende Pixel in a) ist zweifellos mit sehr hoher Wahrscheinlichkeit schwarz, doch welchen Grauwert hat der fehlende Bildpunkt in b)? Offensichtlich gibt es zwei vermutlich gleichwahrwahrscheinliche Favoriten, zwischen denen eine Entscheidung schwerfällt. Faßt man c) und d) als mögliche Erweiterungen des unterliegenden Gitters auf, wird man in c) sicher schwarz den Vorrang geben, wohingegen in d) grau naheliegender zu sein scheint.

Wodurch aber wurden die Vermutungen über den Grauwert des fehlenden Bildpunktes gerechtfertigt? Doch sicher durch intuitive Regelmäßigkeitsannahmen und damit durch die Grauwerte der umliegenden Bildpunkte! Mögen die Beispiele aus Abbildung 2.1 noch konstruiert erscheinen, so sind sicher

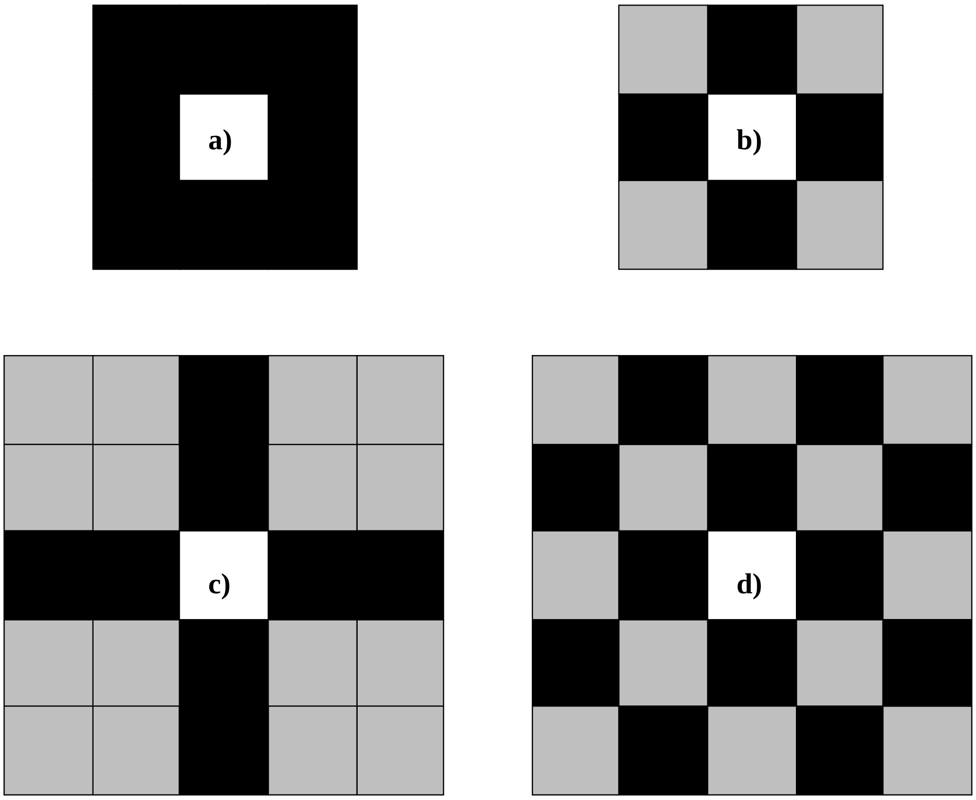


Abbildung 2.1: Vier Pixelbilder mit fehlendem mittleren Bildpunkt

auch in Abbildung 2.2 plausible Vermutungen (Glattheitsannahmen) möglich. Man wähle dazu eine beliebige Bildstelle, denke sich den Grauwert an dieser Stelle weg und versuche ihn nur durch die Grauwerte seiner Umgebung wieder zu begründen.

Es läßt sich somit festhalten, daß es gegenseitige Abhängigkeiten der Grauwerte beieinanderliegender Bildpunkte zu geben scheint, die sich mit der Größe des betrachteten Ausschnitts durchaus ändern können (wie etwa in den Fällen b)–d)). Als nächstes sind diese Abhängigkeiten in das Modell zu integrieren und als zusätzliche Wissensquelle gegenüber der Bildbeschreibung durch Ortsbereichsfunktionen nutzbar zu machen.

2.2 Markoffsche Zufallsfelder

Im vorangegangenen Abschnitt wurden Abhängigkeiten zwischen den Grauwerten benachbarter Pixel als ein charakteristisches Merkmal von Bildern postuliert. Es soll nun ein formaler Beschreibungsapparat entwickelt werden,

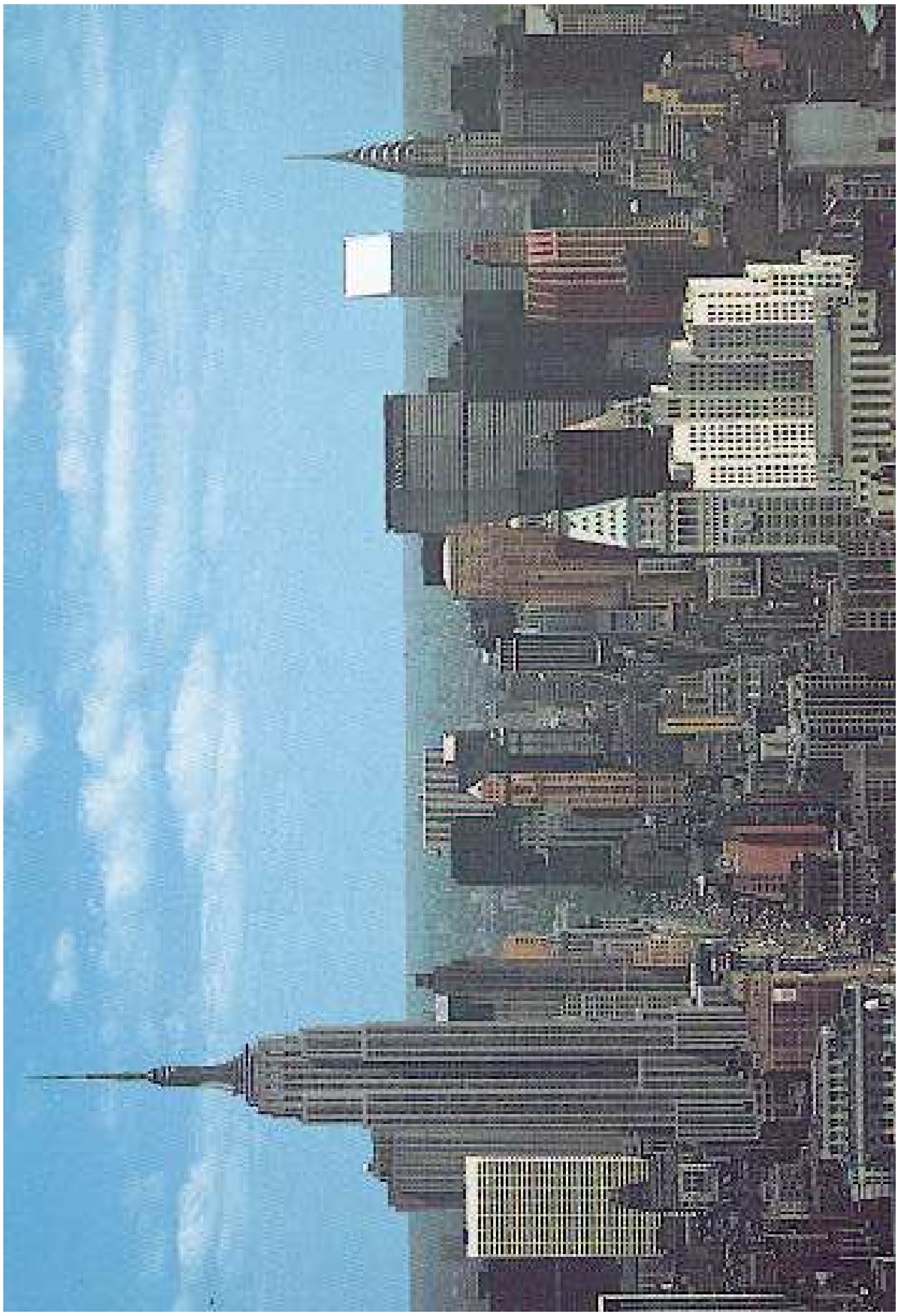


Abbildung 2.2: Blick auf Midtown-Manhattan (Postkartenausschnitt)

der diese Abhängigkeiten modelliert und Grundlage für alle hier folgenden Untersuchungen sein wird.

2.2.1 Nachbarschaft

Zunächst ist festzulegen, was genau unter der Nachbarschaft von Pixeln verstanden werden soll.

Definition 2.4 (Nachbarschaftssystem)
Ein Nachbarschaftssystem *auf einem Gitter L ist eine Menge*

$$\eta = \{\, \eta_{ij} \subset L \,|\, (i,j) \in L \,\}$$

von Teilmengen der Gitterpunkte, die folgende Bedingungen erfüllen:

$$\begin{array}{rll} (i) & (i,j) \notin \eta_{ij} & \forall\, i = 1,\dots,N,\ j = 1,\dots,M \\ (ii) & (i,j) \in \eta_{kl} \iff (k,l) \in \eta_{ij} & \forall\, i,k = 1,\dots,N,\ j,l = 1,\dots,M \end{array}$$

η_{ij} *bezeichnet die* Nachbarschaft *des Punktes $(i,j) \in L$, und mit $\overline{\eta_{ij}}$ wird die* geschlossene Nachbarschaft $\eta_{ij} \cup \{(i,j)\}$ *abgekürzt.*

Definition 2.4 ist eine Formalisierung des allgemeinsten Falls, in der lediglich vernünftige Forderungen an ein System von Nachbarschaften festgeschrieben sind. Es wird beispielsweise keinerlei Aussage darüber gemacht, wie weit benachbarte Pixel auseinanderliegen dürfen, oder ob die Nachbarn eines Pixels eine zusammenhängende Region ergeben müssen. Einige Beispiele sollen verdeutlichen, welche Freiheiten bei der Wahl von η verbleiben.

Beispiel 2.5 (Nachbarschaftssysteme)

(i) *Es gelte $\eta_{ij} = \emptyset$ für alle $(i,j) \in L$. Dann ist η das Nachbarschaftssystem, dessen Nachbarschaften leer sind. Da kein Pixel einen Nachbarn hat, können in einem solchen System auch keine Abhängigkeiten auftreten (vgl. Beispiel 2.3).*

(ii) *Dual zu (i) beschreibt $\eta = \{\, L \setminus \{(i,j)\} \,|\, (i,j) \in L \,\}$ eine vollständige Nachbarschaft, in der jeder Gitterpunkt zu jedem anderen benachbart ist.*

(iii) *Als Beispiel für ein System nicht zusammenhängender Nachbarschaftsmengen betrachte man die Nachbarschaften*

$$\eta_{ij} = \{(k,l) \in L | (i,j) \neq (k,l),\ k \equiv i \bmod 2,\ l \equiv j \bmod 2\} \qquad (2.1)$$

und überprüfe anhand der Definition, daß es sich tatsächlich um ein Nachbarschaftssystem handelt. In Abbildung 2.3 sind in η zueinander benachbarte Pixel gleichfarbig dargestellt.

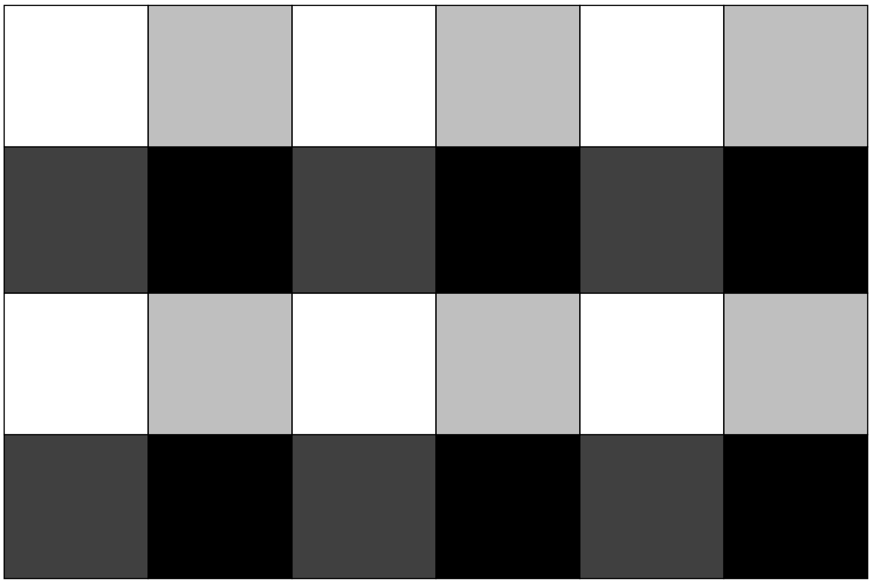


Abbildung 2.3: Periodisches Nachbarschaftssystem

(iv) *Nachbarschaftssysteme können aber ebensogut ein völlig unregelmäßiges Aussehen aufweisen. In Abbildung 2.4 sind wie beim vorherigen Beispiel alle untereinander benachbarten Pixel gleich eingefärbt. Die zugehörigen Nachbarschaften enthalten zu einem gegebenem Pixel alle anderen Pixel desselben Grauwertes.*

(v) *Als wichtigstes Beispiel soll eine hierarchische Folge $\eta^0, \eta^1, \eta^2, \ldots$ von Nachbarschaftssystemen erläutert werden. Dabei enstpricht η^0 dem System leerer Nachbarschaften aus (i), η^1 enthält zu jedem Punkt seine vier nächsten Nachbarn, η^2 die jeweils acht nächsten Nachbarn, usw. Für die Punkte am Rande des Gitters sind entsprechend weniger Nachbarn in den η_{ij} enthalten*[2]*. Für hinreichend große n entspricht η^n dem System aus Beispiel (ii). Abbildung 2.5 zeigt die Nachbarn eines inneren Punktes (i, j) in den Systemen $\eta^1, \ldots, \eta^4$.*

Für die Zwecke der Bildverarbeitung sind insbesondere die Elemente der hierarchischen Folge aus Beispiel 2.5 *(v)* von Bedeutung. Am häufigsten beschränkt man sich jedoch auf η^2, da die Nachbarschaften überschaubar und

[2]Will man dennoch alle η_{ij} gleichmächtig machen, kann man das Gitter als Torus interpretieren und beispielsweise in die Nachbarschaft eines Pixels am rechten Rand die passenden Pixel des linken Randes aufnehmen. In der Bildverarbeitung erscheint diese Konstruktion jedoch unnatürlich, weil man im allgemeinen keine Wechselwirkungen zwischen Pixeln an entgegengesetzten Bildrändern erwarten wird (siehe aber auch das autobinomiale Modell in Abschnitt 3.1.2).

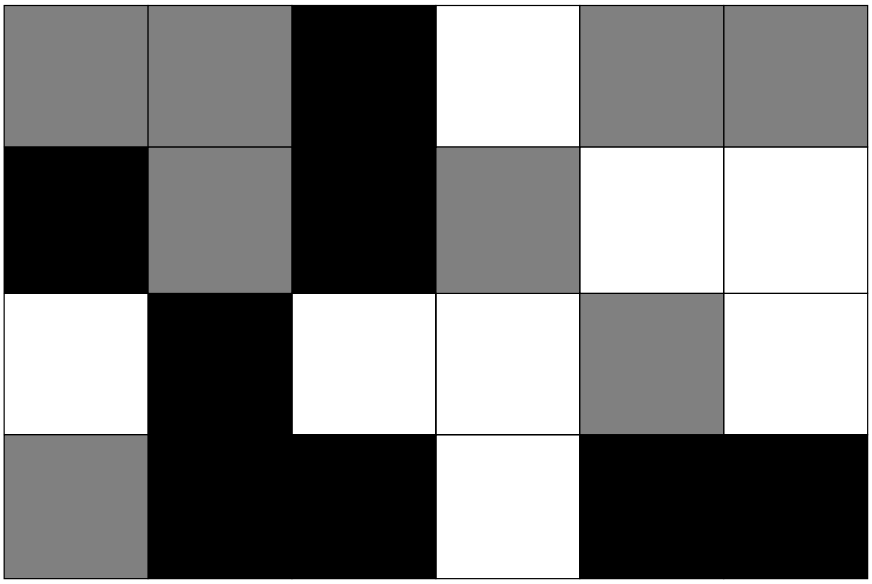


Abbildung 2.4: Unstrukturiertes Nachbarschaftssystem

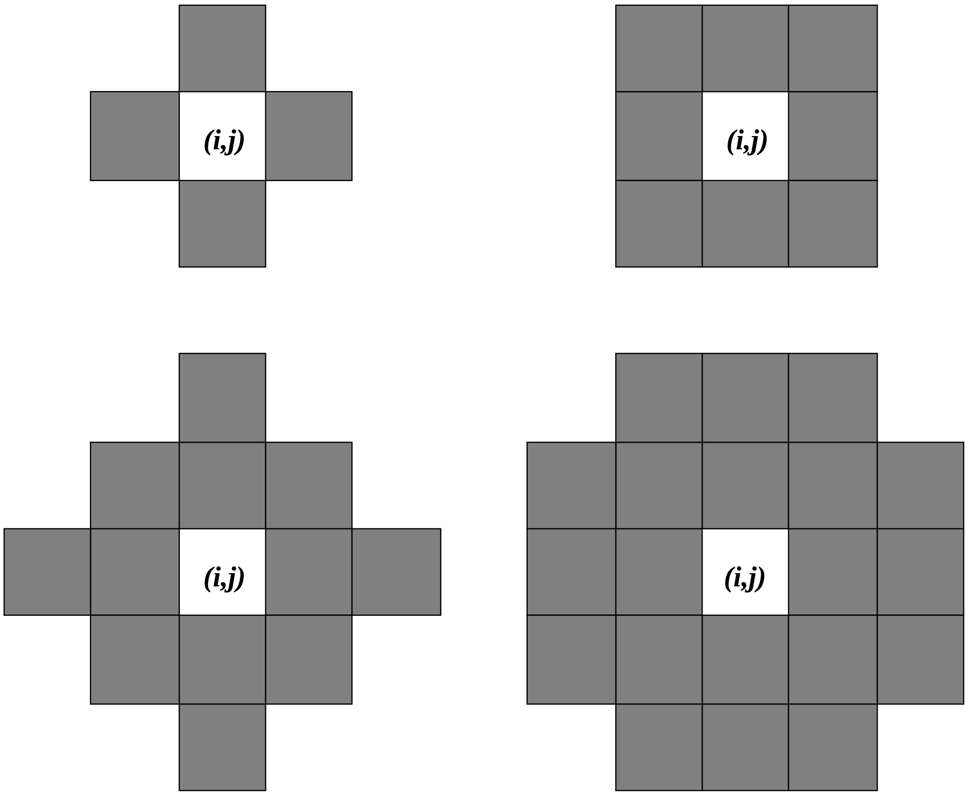


Abbildung 2.5: Nachbarschaften in η^1 bis η^4

rechenerisch handhabbar bleiben, trotzdem aber alle angrenzenden Pixel enthalten. Daß diese Wahl nicht immer angemessen ist, hat bereits Abbildung 2.1 deutlich gemacht.

2.2.2 Markoff-Felder

Nachdem die Vorstellung von Nachbarschaft präzisiert ist, kann nun das eigentliche Ziel dieses Abschnitts, nämlich die Beschreibung von Abhängigkeiten zwischen den Grauwerten der Pixel in einer Nachbarschaft, in Angriff genommen werden.

Man bedient sich dazu der in der Stochastik üblichen Formulierung über bedingte Wahrscheinlichkeiten. Wie bei den vielerorts verwendeten Markoff-Ketten (vgl. z.B. [MP90]) soll die Wahrscheinlichkeit für eine Realisierung einer Zufallsvariable nur von den Realisierungen bestimmter anderer Zufallsvariablen abhängen. Da es im Falle eines Zufallsfeldes, das über einem zweidimensionalen Gitter definiert ist, keine bevorzugte Richtung für eine den Markoff-Ketten entsprechende Kausalität gibt, ist die Erweiterung des Abhängigkeitsgedankens nicht trivial. Trotzdem steht durch die vorstehende Definition von Nachbarschaftssystemen bereits ein ausreichendes Hilfsmittel zur Verfügung, um ihn auf Zufallsfelder zu übertragen.

Im folgenden bezeichne für ein beliebiges Teilgitter $\Lambda \subseteq L$ der Index Λ an einem Vektor dessen Einschränkung auf das Teilgitter Λ.

Definition 2.6 (Markoffsches Zufallsfeld)
Ein Zufallsfeld X über einem Gitter L heißt Markoffsches Zufallsfeld (Markoff-Feld, MRF) *bezüglich eines Nachbarschaftssystems η, falls*

(i) (Positivität)
Für alle $x \in \Gamma^L$ gilt:
$$P(X = x) > 0 \tag{2.2}$$

(ii) (Markoff-Eigenschaft)
Für alle $x = (x_{1,1}, \ldots, x_{N,M}) \in \Gamma^L$ und $(i, j) \in L$ gilt:
$$P\left(X_{ij} = x_{ij} | X_{L\setminus(i,j)} = x_{L\setminus(i,j)}\right) = P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}\right) \tag{2.3}$$

Die Forderung nach positiver Wahrscheinlichkeit für jedes mögliche Bild in *(i)* scheint unbegründet, und für den Moment soll als Rechtfertigung ausreichen, daß auf diese Weise alle bedingten Wahrscheinlichkeiten in *(ii)* wohldefiniert sind. Der eigentliche Grund wird dann in Abschnitt 2.4 deutlich werden. Die rechten Seiten der Formeln (2.3) sollen im folgenden als *lokale Charakteristiken* bezeichnet werden, weil sie für jeden Bildpunkt angeben, in welcher Weise sein Grauwert von denen der Nachbarschaft abhängt.

Für eine präzise Definition der wichtigen Klasse homogener Markoff-Felder wird hier der Begriff der Gittertranslation eingeführt (vgl. Abbildung 2.6):

Definition 2.7 (Gittertranslation)
Es seien L ein Gitter, $d_i, d_j \in \mathbb{Z}$ mit $0 \leq |d_i| \leq N-1$, $0 \leq |d_j| \leq M-1$ und

$$L_{d_i d_j} := \{\,(i,j) \in L \,|\, 1-d_i \leq i \leq N-d_i,\; 1-d_j \leq j \leq M-d_j\,\}\,.$$

Eine (Gitter-)Translation *ist dann eine Abbildung $\tau_{d_i d_j} : L_{d_i d_j} \to L$ mit*

$$\tau_{d_i d_j}(i,j) = (i+d_i, j+d_j) \qquad \forall\,(i,j) \in L_{d_i d_j}.$$

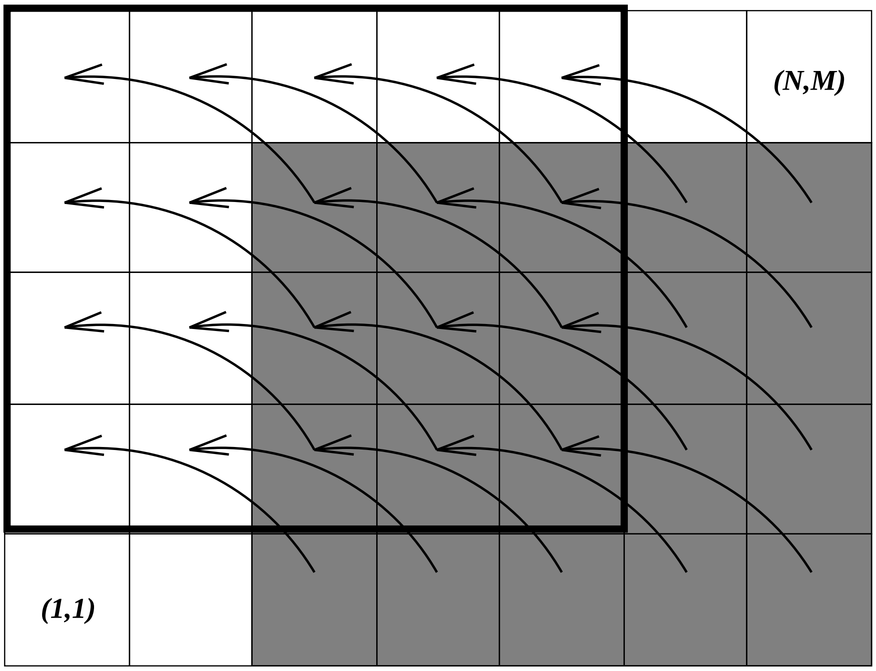


Abbildung 2.6: Gittertranslation

Beispiel für eine Gittertranslation mit $(d_i, d_j) = (-2, 1)$. Der graue Bereich bildet das Teilgitter $L_{d_i d_j}$, der dick umrandete den Wertebereich von $\tau_{d_i d_j}$.

Ein homogenes Markoff-Feld ist nun eines, in dem die lokalen Charakteristiken nur von der Konfiguration der Nachbarschaft abhängen, nicht jedoch von der Lage auf dem Gitter.

Definition 2.8 (Homogenes Markoff-Feld)
Ein Markoff-Feld X über L bezüglich η heißt homogen*, falls für alle Translationen $\tau_{d_i d_j}$ von L und alle $(i,j) \in L_{d_i d_j}$ gilt:*

$$\begin{aligned} & P\left(X_{ij} = x_{ij} \,|\, X_{\eta_{ij} \cap L_{d_i d_j}} = x_{\eta_{ij} \cap L_{d_i d_j}}\right) \\ & = \; P\left(X_{\tau_{d_i d_j}(i,j)} = x_{\tau_{d_i d_j}(i,j)} \;\middle|\; \begin{array}{l} X_{\{\tau_{d_i d_j}(k,l) | (k,l) \in \eta_{ij} \cap L_{d_i d_j}\}} \\ \quad = x_{\{\tau_{d_i d_j}(k,l) | (k,l) \in \eta_{ij} \cap L_{d_i d_j}\}} \end{array}\right) \end{aligned}$$

2.2.3 Anmerkungen

Markoff-Felder sind eine der Intuition nahestehende Bildbeschreibung, in der die in Abschnitt 2.1 postulierten typischen Bildeigenschaften anschaulich ausgedrückt sind. Die Anfänge der Beschreibung nichtkausaler Systeme von Zufallsvariablen über bedingte Wahrscheinlichkeiten liegen (für *nearest neighbor* Systeme) bei Bartlett [Bar55, Abschnitt 2.22][Bar67, Bar68] und vor allem bei Dobrushin [Dob68].

Man muß jedoch kritisch hinterfragen, wie wertvoll ein solches Modell ist. Denn beim Versuch seiner Anwendung ergeben sich erhebliche Schwierigkeiten, die von Derin [Der85] wie folgt zusammengefaßt werden:

- *Konsistenzprobleme:* Die gegebenen lokalen Charakteristiken definieren nicht notwendigerweise eindeutig und konsistent ein Zufallsfeld mit einer gemeinsamen Verteilung der Komponenten.
- *Unbekannte gemeinsame Verteilung:* Die gemeinsame Verteilung der Zufallsvariablen an den Gitterpunkten ist (falls sie überhaupt existiert) aus den lokalen Charakteristiken nicht ohne weiteres ersichtlich.
- *Mangelnde Kausalität:* Anders als bei den Markoff-Ketten gibt es keine eindeutige oder bevorzugte Kausalitätsrichtung.
- *Rechentechnische Probleme:* Die Unkenntnis der gemeinsamen Verteilung und die mangelnde Kausalität erschweren die Verarbeitung von Bilddaten.

Es verwundert daher nicht, daß lange Zeit zahlreiche verwandte Ansätze zur Erweiterung Markoffscher Abhängigkeit von eindimensionalen (zeitindizierten) Systemen auf höherdimensionale (räumliche) Systeme untersucht wurden.

Viele Autoren waren zunächst bemüht, mit dem Abhängigkeitsgedanken auch seine Kausalität aus dem Eindimensionalen in höhere Dimensionen hinüberzuretten, weil sie eine natürliche Grundlage für die iterative oder rekursive Verarbeitung der Zufallsfelder ist [AHK65, JA74, SK76, Woo72].

Der wohl erste dieser Ansätze ist einer der wichtigsten geblieben. Abend *et al.* [AHK65, Kan80] forderten anstelle von (2.3), daß die unter allen weiter "links" oder "unten" gelegenen Komponenten bedingte Verteilung einer Komponente nur von jenen nahe bei dieser gelegenen abhänge, also z.B.

$$\begin{aligned} &P\left(X_{ij} = x_{ij} \,|\, X_{kl} = x_{kl},\, k < i \text{ oder } l < j\right) \\ &\quad = \; P\left(X_{ij} = x_{ij} \,\middle|\, \begin{array}{c} X_{i-1,j} = x_{i-1,j} \\ X_{i-1,j-1} = x_{i-1,j-1} \quad X_{i,j-1} = x_{i,j-1} \end{array}\right). \end{aligned} \tag{2.4}$$

Die so definierten Zufallsfelder werden auch als *Markov Mesh Random Fields* (MMRF) bezeichnet. In Abbildung 2.7 ist leicht zu erkennen, daß es sich bei obigem Beispiel sozusagen um ein *nearest neighbor*-Modell mit Ausbreitungsrichtung von "links nach rechts" und von "oben nach unten" handelt. Die Weitsichtigkeit der Autoren drückt sich unter anderem darin aus, daß sie für beliebige Wahlen von relevanten Vorgängern die Existenz von Nachbarschaften im oben definierten Sinn nachweisen (so daß Beziehungen der Form (2.4) auf solche der Form (2.3) führen). Der volle Zusammenhang, insbesondere mit den Gibbs-Feldern des nächsten Abschnitts, war ihnen jedoch noch unbekannt.

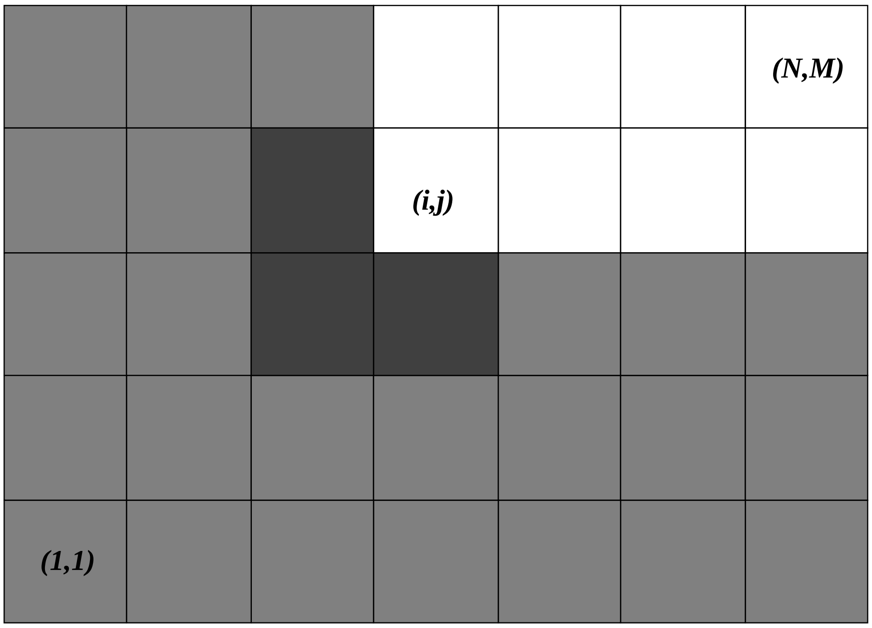


Abbildung 2.7: Kausalität in MMRFs

Die unter allen grauen Pixeln bedingte Wahrscheinlichkeitsverteilung von X_{ij} hängt nur von den Pixeln im dunkelgrauen Bereich ab.

Heute machen diese Ansätze nur noch in maßgeschneiderten Anwendungen Sinn, weil zu Beginn der 70er Jahre der Zusammenhang mit den Gibbs-Feldern, einer durch die Form der gemeinsamen Verteilung der Komponenten definierten Klasse von Zufallsfeldern, entdeckt wurde. Sie sind Gegenstand des nächsten Abschnitts.

2.3 Gibbs-Felder

Im vorangegangenen Abschnitt wurden Abhängigkeiten zwischen den einzelnen Komponenten eines Zufallsfeldes durch bedingte Wahrscheinlichkeiten beschrieben. Es wurde allerdings auch schon darauf hingewiesen, daß die Herleitung der gemeinsamen Verteilung dieser Komponenten erhebliche Schwierigkeiten bereiten kann. Für inkonsistente lokale Charakteristiken existiert sie nicht einmal.

Geht man den umgekehrten Weg, gibt man also eine gemeinsame Verteilung vor, so sollte es im allgemeinen keine nennenswerten Probleme bereiten, daraus bedingte Wahrscheinlichkeiten zu errechnen. Man beachte aber, daß diese nicht notwendig lokale Charakteristiken eines unterstellten Nachbarschaftssystems sind, weil für beliebige gemeinsame Verteilungen die Gleichheit in (2.3) im allgemeinen nicht gilt.

Trotzdem befaßt sich dieser Abschnitt mit einer durch ihre Verteilung bestimmten Klasse von Zufallsfeldern, weil im folgenden Abschnitt gezeigt werden kann, daß sie bezüglich desselben Nachbarschaftssystems äquivalent zur Klasse der Markoffschen Zufallsfelder ist. Da die Charakterisierung durch die gemeinsame Verteilung der Komponenten einer Analogie zur statistischen Physik entspringt, wurden auch die meisten Bezeichungen von dort übernommen. Die Ausführungen bezüglich der physikalischen Analogie sind im wesentlichen in [GNS87] und [Rue68] nachzulesen.

2.3.1 Cliquen und Potentiale

Die statistische Mechanik betrachtet Systeme aus vielen Teilchen (etwa Atomen oder Molekülen), die jeweils bestimmte Zustände einnehmen können. Die makroskopischen ("globalen", meßbaren) Zustandsgrößen werden durch Mittelung über die mikroskopischen Eigenschaften beschrieben, wobei ein Mikrozustand durch die Vielteilchen-Konfiguration der Einteilchen-Zustände gegeben ist.

Die auf ein Teilchen wirkenden Kräfte sind beschrieben durch ein äußeres Kraftfeld und die von den anderen Teilchen auf dieses eine ausgeübten Kräfte. Da zu jeder Kraft eine entsprechende Gegenkraft wirkt, bedingen die im System vorhandenen *Wechselwirkungen* eine symmetrische Nachbarschaftsrelation auf den Teilchen. Werden dann Teilchen als Gitterpunkte interpretiert, so induziert diese Relation auf dem entsprechenden Gitter ein Nachbarschaftssystem. Die Analogie zum eingeführten Bildmodell wird vervollständigt, wenn man die für ein Teilchen möglichen Zustände als mögliche Grauwerte des enstprechenden Gitterpunktes auffaßt.

Da ein Gitter zusammen mit einer (in Form eines Nachbarschaftssystems formulierten) Nachbarschaftsrelation einen Graphen darstellt, kann ein für die weitere Argumentation benötigter Begriff aus der Graphentheorie übernommen werden:

Definition 2.9 (Clique)
Es sei η ein Nachbarschaftssystem über einem Gitter L. Eine nichtleere Teilmenge $C \subseteq L$ heißt Clique *in η, falls für alle $(i,j) \neq (k,l)$ aus C gilt*

$$(k,l) \in \eta_{ij}.$$

Die Menge aller Cliquen in η wird mit $\mathcal{C} = \mathcal{C}(\eta)$ bezeichnet.

Demnach ist eine Menge von Gitterpunkten genau dann eine Clique, wenn ihre Elemente paarweise benachbart sind. Insbesondere ist $\{(i,j)\} \in \mathcal{C}$ für alle $(i,j) \in L$. Abbildung 2.8 zeigt die Cliquentypen der beiden wichtigsten Nachbarschaftsysteme, η^1 und η^2.

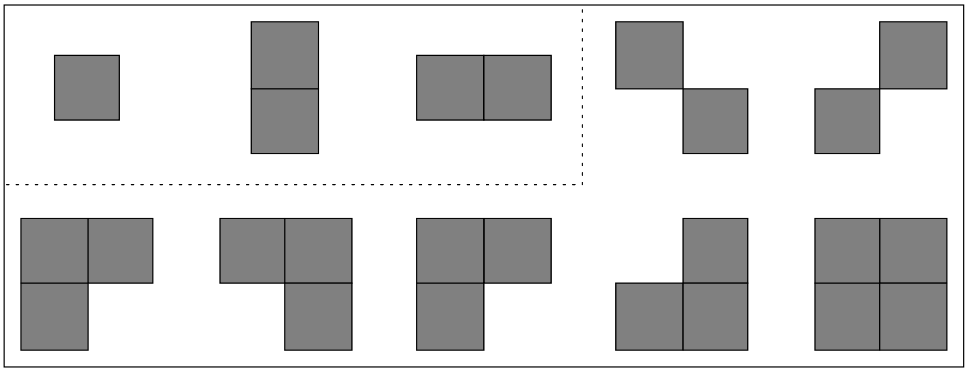


Abbildung 2.8: Cliquentypen in η^1 (gestrichelter Kasten) und η^2 (durchgezogener Kasten)

Da innerhalb einer Clique alle Elemente paarweise benachbart sind, verschwinden die jeweiligen Summen der Wechselwirkungskräfte eines physikalischen Systems, so daß jeder Clique ein Potential zugeordnet werden kann. Bei einelementigen Cliquen ist es durch das äußere Kraftfeld bestimmt.

Eine der zentralen makroskopischen Zustandsgrößen ist die Energie eines Systems. Sie setzt sich zusammen aus der kinetischen und der potentiellen Energie aller Teilchen des Systems und kann daher als Summe aller Cliquenpotentiale ausgedrückt werden.

Definition 2.10 ((Cliquen-)Potential, Energie)
Ein Potential *zu einem Zufallsfeld X über L mit Nachbarschaftssystem η ist eine Schar $V = \left\{V_C : \Gamma^L \to I\!R \,|\, C \in \mathcal{C}(\eta)\right\}$ von Funktionen, wenn aus $x_{ij} = x'_{ij} \quad \forall\, (i,j) \in C$ folgt:*

$$V_C(x) = V_C(x'). \tag{2.5}$$

V_C heißt Cliquenpotential *von C. Gilt für alle $C \in \mathcal{C}$, daß $V_C(x) = 0$, falls $x_{ij} = 0$ für ein $(i,j) \in C$, so heißt V* normalisiert*. Die durch*

$$U(x) = \sum_{C \in \mathcal{C}} V_C(x),$$

definierte Funktion $U = U_V : \Gamma^L \to I\!R$ heißt Energie *zum Potential V.*

Die Definition besagt in Übereinstimmung mit der physikalischen Interpretation, daß das Potential einer Clique lediglich von den Zuständen der in ihr enthaltenen Gitterpunkte abhängt. Ferner ist die Energie einer Realisierung (Konfiguration) des Zufallsfeldes die Summe aller Cliquenpotentiale.

2.3.2 Gibbs-Verteilung

Obwohl für ein abgeschlossenes System bei einem durch die natürlichen Variablen Energie, Volumen und Teilchenzahl gegebenen Makrozustand im allgemeinen noch verschiedene Mikrozustände möglich sind, werden diese nicht weiter unterschieden, da sie nicht beobachtbar sind. Außerdem ist das eigentliche Ziel ja eine sinnvolle Verteilung auf der Menge Γ^L *aller* Mikrozustände, weswegen die Energie nicht fest vorgegeben werden darf.

Es wird daher der Fall eines Systems in Wechselwirkung mit seiner Umgebung betrachtet, bei dem man anstelle der Energie nur die Temperatur $T > 0$ des betrachteten Systems festlegt. Man kannn es sich als in ein Wärmebad getaucht vorstellen, mit dem zusammen es ein abgeschlossenes System von fester Gesamtenergie bildet. Das Bad stellt ein Energiereservoir dar, das idealerweise als unendlich groß angenommen wird. Das eingetauchte System kann dadurch jede mögliche Energie und somit auch jeden möglichen Mikrozustand annehmen. Es bleibt die Frage, mit welcher Wahrscheinlichkeit ein bestimmter Zustand angenommen wird.

Man betrachtet dazu eine Menge von $\mathcal{N}$ identischen Systemen gleicher Teilchenzahl mit vorgegebenem Volumen und vorgegebener Temperatur T, aber nicht fixierter Energie (kanonisches Ensemble). Jedes dieser Systeme befindet sich in einem Mikrozustand $x \in \Gamma^L$, und $n(x)$ sei die Anzahl der Ensembleelemente in diesem Zustand. Natürlich muß dann gelten

$$\mathcal{N} = \sum_{x \in \Gamma^L} n(x). \tag{2.6}$$

Befinden sich die Gesamtsysteme aus Wärmebad und eingetauchtem System jeweils im Gleichgewicht[3], muß sich bei genügend großem $\mathcal{N}$ eine mittlere Energie $\bar{U}$ der Elemente des Ensembles mit

$$\mathcal{N} \cdot \bar{U} = \sum_{x \in \Gamma^L} n(x) \cdot U(x) \tag{2.7}$$

ergeben. Man ist daher an der wahrscheinlichsten Verteilung der Anzahlen auf die Zustände interessiert. Nun sind die Anzahlvektoren n aber multinomialverteilt mit identischen[4] Erfolgswahrscheinlichkeiten $\frac{1}{|\Gamma^L|} = \frac{1}{G^{N \cdot M}}$ für alle Phasenraumpunkte, d.h. die Wahrscheinlichkeit dafür, daß die Mikrozustände x im Ensemble jeweils $n(x)$ mal auftreten, ist

$$\mathcal{N}! \cdot \prod_{x \in \Gamma^L} \frac{\left(\frac{1}{G^{N \cdot M}}\right)^{n(x)}}{n(x)!}. \tag{2.8}$$

Um die wahrscheinlichsten Anzahlen $n^*(x)$ zu erhalten, ist demnach (2.8) unter den Nebenbedingungen (2.6) und (2.7) zu maximieren. Daraus ergibt sich, daß die relativen Häufigkeiten $\frac{n^*(x)}{\mathcal{N}}$ für wachsendes $\mathcal{N}$ gegen

$$\frac{e^{-U(x)/kT}}{\sum_{y \in \Gamma^L} e^{-U(y)/kT}}$$

streben, wobei $k \approx 1.38066 \cdot 10^{-23}$ JK^{-1} die Boltzmann[5]-Konstante bezeichnet, welche sich durch eine weitere, thermodynamisch motivierte Überlegung aus der Maximierung ergibt [GNS87, Seite 167ff]. Die über obige Quotienten definierte Verteilung der Mikrozustände heißt kanonische, Boltzmann- oder Gibbs[6]-Verteilung. Sie ist die gesuchte gemeinsame Verteilung der Komponenten:

[3] Der Gleichgewichtszustand wird definiert als "derjenige makroskopische Zustand eines abgeschlossenen Systems, der sich nach hinreichend langer Wartezeit von selbst einstellt und in dem sich die makroskopischen Zustandsgrößen zeitlich nicht mehr ändern" [GNS87, Seite 5].

[4] Es ist das grundlegende Postulat der statistischen Mechanik, daß alle unter den gegebenen Bedingungen möglichen Mikrozustände gleichberechtigt sind.

[5] L. Boltzmann, 1834–1906, österreichischer Physiker

[6] J. W. Gibbs, 1839–1903, amerikanischer Physiker ungarischer Herkunft

Definition 2.11 (Gibbs-Zufallsfeld)
Es seien V ein Potential und $T > 0$ eine reelle Zahl. Ein Zufallsfeld X ist ein Gibbs-Zufallsfeld (Gibbs-Feld, GRF) *mit Potential V und* Temperatur *T, wenn*

$$P(X = x) = \frac{1}{Z(T)}\, e^{-U(x)/T} \tag{2.9}$$

mit

$$Z(T) = \sum_{x \in \Gamma^L} e^{-U(x)/T}.$$

Die Verteilung (2.9) eines Gibbs-Zufallsfeldes ist demnach gerade eine Gibbs-Verteilung, bei der Boltzmann-Konstante und Temperatur zu einem Parameter zusammengefaßt sind. Auch die Zustandssumme (*partition function*) $Z(T)$, physikalisch eine wichtige Größe, ist hier lediglich als normierende Konstante von Bedeutung. Die Verteilung spiegelt den physikalischen Sachverhalt, daß eine Konfiguration umso wahrscheinlicher wird, je kleiner ihre Energie ist, unmittelbar wider.

Beispiel 2.12 (Ising-Modell)
Das grundlegende Wechselwirkungsmodell findet man bei Ising [Isi25], der damit Phasenübergänge von Ferromagneten erklären wollte. Auch wenn es nicht die gewünschten Reslutate liefern konnte, hat es in der Physik weitreichende Bedeutung gewonnen. Ein Magnet wird darin durch eine Menge von räumlich angeordneten Elementarmagneten beschrieben. Sie nehmen jeweils einen von zwei Werten an, welche die Ausrichtung in bezug auf das umgebende Magnetfeld codieren. Die zwischen den einzelnen Elementarmagneten bestehenden Wechselwirkungen nehmen mit der Entfernung rasch ab, so daß nur die Wechselwirkungen zwischen nächsten Nachbarn berücksichtigt werden.

In der Formulierung als Zufallsfeld entspricht das Ising-Modell einem Gibbs-Feld über einem Gitter L mit Nachbarschaftssystem η^1 und "Grauwertmenge" $\Gamma = \{-1, 1\}$. Einelementige Cliquen erhalten das Potential 0, zweielementige werden mit

$$V_{\{(i,j),(k,l)\}} = \begin{cases} -I & x_{ij} = x_{kl} \\ I & x_{ij} \neq x_{kl} \end{cases}$$

bewertet, wobei I eine Konstante ist. Das heißt also, daß parallel ausgerichtete Paare von Elementarmagneten die Energie $-I$, antiparallele die Energie $+I$ erhalten. Die Energie eines Magneten in der Konfiguration x berechnet sich dann zu

$$U(x) = -I \sum_{\substack{\{(i,j),(k,l)\} \in C \\ (i,j) \neq (k,l)}} \delta_{x_{ij}, x_{kl}}$$

mit dem Kronecker-Symbol $\delta_{x_{ij},x_{kl}}$, weswegen die Wahrscheinlichkeit dafür, daß der Magnet sich bei Temperatur T in einer Konfiguration x befindet,

$$P(X = x) = \frac{1}{Z(T)}\, e^{-U(x)/T}$$

ist.

Der mit Temperatur bezeichnete Parameter T eines Gibbs-Feldes nimmt Einfluß darauf, wie stark die Wahrscheinlichkeit einer Konfiguration von ihrer Energie abhängt. Für große T, also hohe Temperaturen, ist der Einfluß der Energie gering, d.h. (2.9) ähnelt eher einer Gleichverteilung. Für niedrige Temperaturen jedoch, besonders wenn $T < 1$, werden Zustände hoher Energie sehr unwahrscheinlich.

Potentiale beschreiben somit die Wechselwirkungseigenschaften einer ganzen Klasse von Zufallsfeldern unterschiedlicher Temperatur. Diese Eigenschaften sind in den Elementen dieser Klasse – abhängig von T – unterschiedlich stark ausgeprägt bzw. verwischt. Unterscheiden sich zwei Potentiale nur durch einen festen Faktor c, so lassen sie sich durch Anpassung der Temperatur ineinander überführen und bestimmen daher dieselbe Klasse von Zufallsfeldern. Man verifiziere außerdem, daß zwei Potentiale, deren Cliquenpotentiale sich lediglich um einen festen Betrag d unterscheiden, die gleiche Verteilung definieren.

Von besonderer Bedeutung ist der Grenzübergang $T \to 0$, weil die Gibbs-Verteilung dann sogar gegen eine Gleichverteilung auf den Zuständen minimaler Energie des durch das Potential bestimmten Zufallsfeldes konvergiert. In Kapitel 4 wird darauf zurückgekommen, und in 4.2.3 ein Nachweis für diese Behauptung geführt.

Gelegentlich wird auf die Angabe der Temperatur verzichtet. In diesen Fällen wird stillschweigend $T = 1$ angenommen.

Ein starker Befürworter der globalen Betrachtungsweise von Zufallsfeldern war Whittle [Whi63]. Aufgrund des Zusammenhangs mit der statistischen Mechanik sah er in der gemeinsamen Verteilung der Komponenten eine "natürlichere" Beschreibung von Zufallsfeldern als es die bedingten Wahrscheinlichkeiten sind. Er schlug vor, Zufallsfelder grundsätzlich so zu spezifizieren, daß ihre gemeinsame Verteilung durch das Produkt von auf Teilmengen des Gitters definierten Funktionen gegeben ist, was sich nun in der Summe der Cliquenpotentiale im Exponenten von (2.9) widerspiegelt.

Mittlerweile allerdings hat sich die Frage erübrigt, welcher der beiden Ansätze zum "besseren" Modell führt, denn wie im nächsten Abschnitt gezeigt wird, sind beide äquivalent. Folglich ist es mit Hilfe der Gibbs-Verteilung möglich, die in 2.2 erwähnten Schwachstellen der Markoff-Felder zu kompensieren.

2.4 Äquivalenz der Charakterisierungen

Die globalen Eigenschaften von Gibbs-Zufallsfeldern sind erheblich leichter zu überblicken als die der Markoffschen, da jene eben gerade über ihre gemeinsame Verteilung (2.9) definiert sind. Andererseits sind lokale Eigenschaften in den bedingten Wahrscheinlichkeiten (2.3) offensichtlicher ausgedrückt. Es war daher ein echter Durchbruch, als Anfang der 70er Jahre Spitzer [Spi71] und Averintzev [Ave70] unabhängig voneinander die Äquivalenz von bestimmten Markoff- und Gibbs-Feldern bezüglich desselben Nachbarschaftssystems nachwiesen.

Inzwischen wurden mehrere Beweise der vollen Äquivalenz gegeben. Der erste davon geht auf Hammersley und Clifford zurück, wurde aber von ihnen nicht veröffentlicht, weil sie zuvor die Positivitäts-Bedingung (2.2) eliminieren wollten (was sich später als unmöglich herausstellte [Bes74, angeschlossene Diskussion]).

Die Äquivalenz der Charakterisierungen soll in diesem Abschnitt ausführlich bewiesen werden, weil durch sie die in Abschnitt 2.2.3 aufgeführten Nachteile von Markoffschen Zufallsfeldern überwunden werden.

Der Beweis erfolgt in zwei Schritten, aus denen sich explizite Formeln zur Umrechnung von Cliquenpotentialen in lokale Charakteristiken und umgekehrt ergeben werden. Zunächst wird die einfachere Richtung gezeigt werden, nämlich daß jedes Gibbs-Feld bezüglich η auch ein Markoff-Feld bezüglich desselben η ist. Die Benutzung der Möbius-Inversionsformel für den Beweis der Umkehrung findet sich zuerst bei Grimmett [Grim72], die hier vorgestellte Fassung folgt allerdings weitgehend dem von Griffeath in [Grif76] geführten, weil dort die lokalen Charakteristiken in den Umrechnungsformeln deutlicher hervortreten.

Es ist hilfreich, zuvor eine abkürzende Schreibweise für ein durch Änderung von Einträgen aus x hervorgehendes Bild einzuführen. Für $\Lambda \subseteq L$ und Grauwerte $y_{kl} \in \Gamma$, $(k,l) \in \Lambda$, sei

$$x_{ij}^{y_\Lambda} = x_{ij}^{\{y_{kl}|(k,l)\in\Lambda\}} := \begin{cases} y_{ij} & \text{falls } (i,j) \in \Lambda \\ x_{ij} & \text{sonst} \end{cases}$$

und

$$x^{y_\Lambda} := (x_{1,1}^{y_\Lambda}, \dots, x_{N,M}^{y_\Lambda}) \tag{2.10}$$

das Bild x, dessen Grauwerte an den Stellen in Λ ersetzt wurden. Für Änderungen $x^{y_{\{(i,j)\}}} = x^{\{y_{ij}\}}$ an einem einzigen Punkt wird abkürzend $x^{y_{ij}}$ geschrieben.

Satz 2.13
Jedes Gibbs-Feld X bezüglich eines Nachbarschaftssystems η ist ein Markoff-Feld bezüglich desselben Nachbarschaftssystems. Für die lokalen Charakteristiken von X gilt:

$$P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}\right) = \frac{1}{Z_{ij}(T)} e^{-\sum\limits_{C \in \mathcal{C}:(i,j) \in C} V_C(x)/T} \tag{2.11}$$

mit

$$Z_{ij}(T) := \sum_{y_{ij} \in \Gamma} e^{-\sum\limits_{C \in \mathcal{C}:(i,j) \in C} V_C(x^{y_{ij}})/T}.$$

Beweis: Zu zeigen sind (i) die Positivität und (ii) die Markoff-Eigenschaft des Zufallsfeldes X. Wegen (2.9) ist (i) offensichtlich, und aus den Rechenregeln für bedingte Wahrscheinlichkeiten folgt für beliebige $x = (x_{1,1}, \ldots, x_{N,M}) \in \Gamma^L$ und $(i,j) \in L$:

$$\begin{aligned}
P\left(X_{ij} = x_{ij} | X_{L\setminus(i,j)} = x_{L\setminus(i,j)}\right) &= \frac{P(X=x)}{P\left(X_{L\setminus(i,j)} = x_{L\setminus(i,j)}\right)} \\
&= \frac{Z(T)^{-1} \exp\{-U(x)/T\}}{Z(T)^{-1} \sum\limits_{y_{ij}\in\Gamma} \exp\{-U(x^{y_{ij}})/T\}} = \frac{\exp\{-U(x)/T\}}{\sum\limits_{y_{ij}\in\Gamma} \exp\{-U(x^{y_{ij}})/T\}} \\
&= \frac{\exp\left\{-\sum\limits_{C\in\mathcal{C}} V_C(x)/T\right\}}{\sum\limits_{y_{ij}\in\Gamma} \exp\left\{-\sum\limits_{C\in\mathcal{C}} V_C(x^{y_{ij}})/T\right\}} \\
&\underset{(*)}{=} \frac{\exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\notin C} V_C(x)/T\right\} \cdot \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x)/T\right\}}{\exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\notin C} V_C(x)/T\right\} \cdot \sum\limits_{y_{ij}\in\Gamma} \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x^{y_{ij}})/T\right\}} \\
&= \frac{\exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x)/T\right\}}{\sum\limits_{y_{ij}\in\Gamma} \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x^{y_{ij}})/T\right\}} \qquad (2.12) \\
&= \frac{\exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x)/T\right\}}{\sum\limits_{y_{ij}\in\Gamma} \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\in C} V_C(x^{y_{ij}})/T\right\}}
\end{aligned}$$

$$
\cdot \frac{\sum\limits_{\substack{y_{kl}\in\Gamma \\ (k,l)\notin\overline{\eta_{ij}}}} \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\notin C} V_C(x^{y_{L\setminus\overline{\eta_{ij}}}})/T\right\}}{\sum\limits_{\substack{y_{kl}\in\Gamma \\ (k,l)\notin\overline{\eta_{ij}}}} \exp\left\{-\sum\limits_{C\in\mathcal{C}:(i,j)\notin C} V_C(x^{y_{L\setminus\overline{\eta_{ij}}}})/T\right\}}
$$

$$
\underset{(*)}{=} \frac{\sum\limits_{\substack{y_{kl}\in\Gamma \\ (k,l)\notin\overline{\eta_{ij}}}} \exp\left\{-\sum\limits_{C\in\mathcal{C}} V_C(x^{y_{L\setminus\overline{\eta_{ij}}}})/T\right\}}{\sum\limits_{\substack{y_{kl}\in\Gamma \\ (k,l)\notin\eta_{ij}}} \exp\left\{-\sum\limits_{C\in\mathcal{C}} V_C(x^{y_{L\setminus\eta_{ij}}})/T\right\}}
$$

$$
= \frac{P\left(X_{\overline{\eta_{ij}}} = x_{\overline{\eta_{ij}}}\right)}{P\left(X_{\eta_{ij}} = x_{\eta_{ij}}\right)} = P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}\right)
$$

Somit gilt für X die Markoff-Eigenschaft (2.3). Man beachte für die mit $(*)$ gekennzeichneten Gleichheiten, daß für alle $C \in \mathcal{C}$ mit $(i,j) \in C$ gilt: $V_C(x) = V_C(x^{y_\Lambda})$ für beliebige Änderungen $y_\Lambda \in \Gamma^\Lambda$, $\Lambda \subseteq (L \setminus \overline{\eta_{ij}})$, außerhalb der geschlossenen Nachbarschaft $\overline{\eta_{ij}}$. Beziehung (2.11) ergibt sich unmittelbar aus (2.12). □

Für die Umkehrung, daß jedes Markoffsche auch ein Gibbs-Zufallsfeld ist, wird eine aus der Kombinatorik bekannte Formel benötigt (vgl. beispielsweise [OHW87]):

Lemma 2.14 (Umkehrformel von Möbius[7])
Für zwei reellwertige, auf den Teilmengen eines Gitters L definierte Funktionen $f, g : \mathcal{P}(L) \to \mathbb{R}$ und jedes $C \in \mathcal{P}(L)$ gilt

$$
g(C) = \sum_{B\subseteq C} f(B) \iff f(C) = \sum_{B\subseteq C} \mu(B,C) g(B),
$$

wobei die Möbiusfunktion $\mu : \mathcal{P}(L) \times \mathcal{P}(L) \to \{-1, 0, 1\}$ *durch*

$$
\mu(B,C) := \begin{cases} (-1)^{|C\setminus B|} & \textit{falls } B \subseteq C \\ 0 & \textit{sonst.} \end{cases}
$$

definiert ist.

Um nun nachzuweisen, daß jedes über die bedingten Wahrscheinlichkeiten der Form (2.3) definierte Markoff-Feld eine gemeinsame Verteilung gemäß (2.9) hat, sei vorab an die Schreibweise (2.10) für geänderte Bilder erinnert sowie daran, daß für Markoffsche Zufallsfelder die Positivitätsbedingung (2.2) gilt. Ferner sei $\mathbf{0} \in \Gamma^L$ das Bild, dessen Komponenten allesamt den Grauwert 0 haben.

[7] A. F. Möbius, 1790–1868, deutscher Mathematiker und Astronom

Satz 2.15
Jedes Markoff-Feld X bezüglich eines Nachbarschaftssystems η ist ein Gibbs-Feld bezüglich desselben Nachbarschaftssystems. Ein normalisiertes Potential des Gibbs-Feldes ist durch

$$V_C(x) = -\sum_{B \subseteq C \setminus \{(i,j)\}} (-1)^{|C \setminus B|} \cdot \log \frac{P\left(X_{ij} = 0 | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)}{P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)} \tag{2.13}$$

für alle $C \in \mathcal{C}$ gegeben.

Beweis: Es ist zunächst zu zeigen, daß die Verteilung von X eine Gibbs-Verteilung ist. Dazu sei $x \in \Gamma^L$ ein beliebiges, aber festes Bild. Dann kann zu jeder Teilmenge $C \subseteq L$ definiert werden:

$$g(C) := \log \frac{P(X = \mathbf{0}^{x_C})}{P(X = \mathbf{0})}$$

$$f(C) := -V_C(x) := \begin{cases} 0 & \text{falls } C = \emptyset \\ \sum\limits_{B \subseteq C} (-1)^{|C \setminus B|} \cdot \log P\left(X = \mathbf{0}^{x_B}\right) & \text{sonst.} \end{cases}$$

Offensichtlich erfüllen die so definierten Potentiale V_C die Bedingung (2.5), man beachte aber, daß ihre Gleichheit mit denen aus (2.13) noch zu zeigen ist. Für $C = \emptyset$ ergibt sich $f(\emptyset) = -V_\emptyset(x) = 0 = \log \frac{P(X=\mathbf{0})}{P(X=\mathbf{0})} = \log \frac{P(X=\mathbf{0}^{\emptyset})}{P(X=\mathbf{0})} = g(\emptyset) = \mu(\emptyset, \emptyset) g(\emptyset)$. Ist C nichtleer, so gilt $\sum_{B \subseteq C} \mu(B, C) = 0$ und folglich

$$\begin{aligned} f(C) &= -V_C(x) = \sum_{B \subseteq C} (-1)^{|C \setminus B|} \log P\left(X = \mathbf{0}^{x_B}\right) \\ &= \left[\sum_{B \subseteq C} \mu(B, C) \log P\left(X = \mathbf{0}^{x_B}\right)\right] - \left[\sum_{B \subseteq C} \mu(B, C)\right] \cdot \log P\left(X = \mathbf{0}\right) \\ &= \sum_{B \subseteq C} \mu(B, C) \log \frac{P\left(X = \mathbf{0}^{x_B}\right)}{P\left(X = \mathbf{0}\right)} \\ &= \sum_{B \subseteq C} \mu(B, C) g(B). \end{aligned}$$

Mit Lemma 2.14 ist daher $\sum_{C \subseteq L} f(C) = g(L)$, und man erhält

$$\begin{aligned} -U(x) &:= -\sum_{C \in \mathcal{C}} V_C(x) = \sum_{C \subseteq L} -V_C(x) \\ &= \sum_{C \subseteq L} f(C) = g(L) \\ &= \log \frac{P(X = \mathbf{0}^{x_L})}{P(X = \mathbf{0})} \\ &= \log \frac{P(X = x)}{P(X = \mathbf{0})}, \end{aligned}$$

wobei schon die erst weiter unten gezeigte Behauptung benutzt wurde, daß $V_C(x) = 0$ für $C \notin \mathcal{C}$. Durch einfache Umformung ergibt sich $P(X = x) = P(X = \mathbf{0}) \cdot e^{-U(x)}$. Setzt man nun $T := 1$ und summiert beide Seiten über alle $x \in \Gamma^L$, so folgt weiterhin $P(X = \mathbf{0})^{-1} = \sum_{x \in \Gamma^L} e^{-U(x)/T} =: Z(T)$, und es gilt daher

$$P(X = x) = \frac{1}{Z(T)}\, e^{-U(x)/T}.$$

Somit ist nur noch zu zeigen, daß V normalisiert ist, nur auf Mengen aus $\mathcal{C}$ nicht verschwindet und die Cliquenpotentiale wie in (2.13) ausgedrückt werden können. Es gilt aber für jedes $C \notin \mathcal{C}$, $C \neq \emptyset$, und beliebige $(i,j) \in C$, $(i',j') \notin \overline{\eta_{ij}}$

$$\begin{aligned}
V_C(x) &= -\sum_{B \subseteq C} (-1)^{|C \setminus B|} \log P\left(X = \mathbf{0}^{x_B}\right) \\
&= -\sum_{B \subseteq C \setminus \{(i,j),(i',j')\}} (-1)^{|C \setminus B|} \Big[\log P\left(X = \mathbf{0}^{x_B}\right) - \log P\left(X = \mathbf{0}^{x_{B \cup \{(i,j)\}}}\right) \\
&\qquad - \log P\left(X = \mathbf{0}^{x_{B \cup \{(i',j')\}}}\right) + \log P\left(X = \mathbf{0}^{x_{B \cup \{(i,j),(i',j')\}}}\right)\Big] \\
&= -\sum_{B \subseteq C \setminus \{(i,j),(i',j')\}} (-1)^{|C \setminus B|} \log \frac{P\left(X = \mathbf{0}^{x_B}\right) P\left(X = \mathbf{0}^{x_{B \cup \{(i,j),(i',j')\}}}\right)}{P\left(X = \mathbf{0}^{x_{B \cup \{(i,j)\}}}\right) P\left(X = \mathbf{0}^{x_{B \cup \{(i',j')\}}}\right)} \\
&= -\sum_{B \subseteq C \setminus \{(i,j),(i',j')\}} (-1)^{|C \setminus B|} \log \left(\frac{P\left(X_{ij} = 0 | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_B}_{L \setminus \{(i,j)\}}\right)}{P\left(X_{ij} = x_{ij} | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_{B \cup \{(i',j')\}}}_{L \setminus \{(i,j)\}}\right)} \right. \\
&\qquad \left. \cdot \frac{P\left(X_{ij} = 0 | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_{B \cup \{(i',j')\}}}_{L \setminus \{(i,j)\}}\right)}{P\left(X_{ij} = x_{ij} | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_B}_{L \setminus \{(i,j)\}}\right)} \right) \\
&= -\sum_{B \subseteq C \setminus \{(i,j),(i',j')\}} (-1)^{|C \setminus B|} \log \left(\frac{P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = \mathbf{0}^{x_{B \cup \{(i',j')\}}}_{\eta_{ij}}\right)}{P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)} \right. \\
&\qquad \left. \cdot \frac{P\left(X_{ij} = 0 | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)}{P\left(X_{ij} = 0 | X_{\eta_{ij}} = \mathbf{0}^{x_{B \cup \{(i',j')\}}}_{\eta_{ij}}\right)} \right) \\
&= 0,
\end{aligned}$$

weil $(i',j') \notin \overline{\eta_{ij}}$. Ist dagegen $C \in \mathcal{C}$, dann gilt für jedes $(i,j) \in C$

$$\begin{aligned}
V_C(x) &= -\sum_{B \subseteq C} (-1)^{|C \setminus B|} \log P\left(X = \mathbf{0}^{x_B}\right) \\
&= -\sum_{B \subseteq C \setminus \{(i,j)\}} (-1)^{|C \setminus B|} \left[\log P\left(X = \mathbf{0}^{x_B}\right) - \log P\left(X = \mathbf{0}^{x_{B \cup \{(i,j)\}}}\right)\right] \\
&= -\sum_{B \subseteq C \setminus \{(i,j)\}} (-1)^{|C \setminus B|} \log \frac{P\left(X = \mathbf{0}^{x_B}\right)}{P\left(X = \mathbf{0}^{x_{B \cup \{(i,j)\}}}\right)} \\
&= -\sum_{B \subseteq C \setminus \{(i,j)\}} (-1)^{|C \setminus B|} \log \frac{P\left(X_{ij} = 0 | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_B}_{L \setminus \{(i,j)\}}\right)}{P\left(X_{ij} = x_{ij} | X_{L \setminus \{(i,j)\}} = \mathbf{0}^{x_B}_{L \setminus \{(i,j)\}}\right)}
\end{aligned}$$

$$= \; - \sum_{B \subseteq C \setminus \{(i,j)\}} (-1)^{|C \setminus B|} \log \frac{P\left(X_{ij} = 0 | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)}{P\left(X_{ij} = x_{ij} | X_{\eta_{ij}} = \mathbf{0}^{x_B}_{\eta_{ij}}\right)}.$$

Damit ist (2.13) gezeigt, und man erkennt sofort, daß $V = \{V_C | C \in \mathcal{C}\}$ ein normalisiertes Potential ist (wähle für obige Darstellung jenes $(i,j) \in C$, für das $x_{ij} = 0$). □

Das Potential aus dem letzten Beweis wird auch als *kanonisches Potential* bezeichnet. An der Tatsache, daß es normalisiert ist, läßt sich im nachhinein noch einmal anschaulicher ablesen, wodurch die Gleichheit von $Z(T)$ und $P(X = \mathbf{0})^{-1}$ zustande kommt. Es ist nämlich $P(X = \mathbf{0}) = e^{-U(\mathbf{0})/T}/Z(T) = e^0/Z(T) = Z(T)^{-1}$.

Zusammengenommen liefern die Sätze 2.13 und 2.15 die gewünschte Äquivalenz.

Korollar 2.16 (Äquivalenz der Charakterisierungen)
Ein Zufallsfeld X auf einem Gitter L ist ein Markoff-Feld bezüglich des Nachbarschaftssystems η genau dann, wenn X ein Gibbs-Feld bezüglich η ist. Lokale Charakteristiken und Cliquenpotentiale lassen sich gemäß (2.11) und (2.13) ineinander überführen.

Durch diese Äquivalenz ist es (mit etwas Rechenaufwand) möglich, ein und dasselbe Zufallsfeld sowohl anhand seiner lokalen Eigenschaften (der lokalen Charakteristiken (2.3)), als auch anhand seiner globalen Eigenschaften (der Verteilung (2.9)) zu beschreiben. Korollar 2.16 stellt mit der Umrechnungsformel (2.13) darüber hinaus einen möglichen Konsistenztest für lokale Charakteristiken bereit. Von größerer Bedeutung ist jedoch die Formel (2.11), weil es im allgemeinen leichter ist, Annahmen oder Vorwissen über ein Zufallsfeld in Cliquenpotentialen auszudrücken. Die daraus abgeleiteten lokalen Charakteristiken erleichtern weitere Berechnungen, weil nur noch die Grauwerte innerhalb einer Nachbarschaft dazu herangezogen werden müssen. Die Anwendungen in Kapitel 3 liefern Beispiele hierfür.

Insgesamt wurde in den ersten Abschnitten dieses Kapitels mit den Zufallsfeldern ein stochastisches Modell für bereits digitalisierte und quantisierte Bilder vorgestellt und gezeigt, daß es für jede durch ihr zugehöriges Nachbarschaftssystem bestimmte Teilklasse dieser Zufallsfelder zwei sehr unterschiedliche Charakterisierungen gibt. Während das Markoffsche Zufallsfeld der intuitivere Ansatz zur Beschreibung von lokalen Abhängigkeiten ist, liefert das Gibbs-Feld eine physikalische Interpretation.

2.5 Einführung einer parametrisierten Teilklasse

In diesem Abschnitt werden spezielle Formen, welche die Energie eines Gibbs-Zufallsfeldes annehmen kann, untersucht. Ausgehend von frühen Modellen wird der Begriff der Parametrisierung präzisiert und eine neue parametrisierte Teilklasse eingeführt, deren Cliquenpotentiale eine sehr anschauliche Interpretation besitzen. Anhand von Beispielen wird weiterhin gezeigt, daß diese Klasse zahlreiche von anderen Autoren vorgestellte verallgemeinert.

2.5.1 Die Hammersley-Clifford-Expansion und Auto-Modelle

Wie bereits im vorherigen Abschnitt erwähnt, gelang es zuerst Hammersley und Clifford, die Äquivalenz beliebiger Markoff- und Gibbs-Felder zu zeigen, indem sie die allgemeinste Form der Verteilung eines Markoff-Feldes angaben.

Der von ihnen geführte Beweis beruht auf der nun so genannten Hammersley-Clifford-Expansion von

$$Q(x) := \log \frac{P(X = x)}{P(X = \mathbf{0})}. \tag{2.14}$$

Diese Expansion ist eindeutig und kann in der Form

$$\begin{aligned} Q(x) &= \sum_{k=1}^{N\cdot M} \sum_{\{(i_1,j_1),\ldots,(i_k,j_k)\}\subseteq L} x_{i_1j_1}\cdots x_{i_kj_k} G_{i_1j_1,\ldots,i_kj_k}(x_{i_1j_1},\ldots,x_{i_kj_k}) \qquad (2.15)\\ &= \sum_{(i,j)\in L} x_{ij}G_{ij}(x_{ij}) \\ &\quad + \sum_{\{(i_1,j_1),(i_2,j_2)\}\subset L} x_{i_1j_1}x_{i_2j_2}G_{i_1j_1,i_2j_2}(x_{i_1j_1},x_{i_2j_2}) + \\ &\quad \vdots \\ &\quad + \sum_{\{(i_1,j_1),\ldots,(i_{N\cdot M},j_{N\cdot M})\}\subseteq L} x_{i_1j_1}\cdots x_{i_{N\cdot M}j_{N\cdot M}} G_{i_1j_1,\ldots,i_{N\cdot M}j_{N\cdot M}}(x_{i_1j_1},\ldots,x_{i_{N\cdot M}j_{N\cdot M}}) \end{aligned}$$

notiert werden. Das Ergebnis des Theorems von Hammersley und Clifford lautet nun, daß es zu jedem Markoff-Feld, für das aufgrund von (2.14) die Positivitätsbedingung gelten muß, G-Funktionen gibt, welche eine Darstellung (2.15) erlauben, und daß es umgekehrt für jede zulässige Wahl der G-Funktionen ein Markoff-Feld mit durch Q bestimmter Verteilung P^X gibt.

Zulässig ist eine Wahl von G-Funktionen, wenn $G_{\{(i_1,j_1),\ldots,(i_k,j_k)\}}$ nur für *Cliquen* $\{(i_1,j_1),\ldots,(i_k,j_k)\}$ ungleich Null ist. Ein eleganter Beweis wird von Besag [Bes74] gegeben.

Der durch den Äquivalenzbeweis des vorausgegangenen Abschnitts geschulte Blick erkennt nun auch leicht, daß es sich bei $Q(x)$ um die Energiefunktion eines normalisierten Potentials mit Cliquenpotentialen

$$V_{\{(i_1,j_1),\ldots,(i_k,j_k)\}} = -x_{i_1j_1}\cdots x_{i_kj_k}G_{i_1j_1,\ldots,i_kj_k}(x_{i_1j_1},\ldots,x_{i_kj_k})$$

handelt. Die freie Wahl der G-Funktionen korrespondiert unmittelbar mit der freien Wahl der Cliquenpotentiale von Gibbs-Zufallsfeldern. Für Zufallsfelder mit einer Verteilung, die durch

$$Q(x) = \sum_{(i,j)\in L} x_{ij}G_{ij}(x_{ij}) + \sum_{(i,j)\neq(k,l)\in L} \beta_{ij,kl}x_{ij}x_{kl},$$

also Cliquenpotentiale der Form

$$\begin{aligned} V_{\{(i,j)\}} &= -x_{ij}G_{ij}(x_{ij}) \\ V_{\{(i,j),(k,l)\}} &= -\beta_{ij,kl}x_{ij}x_{kl} \qquad (i,j)\neq(k,l) \end{aligned}$$

bestimmt ist, prägte Besag [Bes74] den Begriff *Auto-Modelle*. Für die Parameter $\beta_{ij,kl}$ muß grundsätzlich zumindest $\beta_{ij,kl}\equiv\beta_{kl,ij}$ und $\beta_{ij,kl}=0$, falls $(k,l)\notin\eta_{ij}$, gelten. Auto-Modelle berücksichtigen offensichtlich nur Wechselwirkungen zwischen jeweils höchstens zwei Pixeln. Zwar können diese, abhängig vom Nachbarschaftssystem, noch weit auseinanderliegen, Potentiale von drei- oder mehrelementigen Cliquen sind jedoch allesamt identisch Null.

Beispiel 2.17 (Spezielle Auto-Modelle)
Für ein Feld aus binären Zufallsvariablen reduzieren sich die G-Funktionen auf einzelne Parameter, da sie nur dann einen Beitrag zu $Q(x)$ liefern, wenn alle Argumente 1 sind. Dies führt auf Zufallsfelder, deren Q-Funktion die Form

$$Q(x) = \sum_{(i,j)\in L} \alpha_{ij}x_{ij} + \sum_{(i,j)\neq(k,l)\in L} \beta_{ij,kl}x_{ij}x_{kl}$$

annimmt. Für $\vartheta = \alpha_{ij} + \sum_{(k,l)\in\eta_{ij}}\beta_{ij,kl}x_{kl}$ folgt, daß die lokalen Charakteristiken eines solchen Zufallsfeldes einem Bernoulli-Experiment mit "Erfolgswahrscheinlichkeit" (Wahrscheinlichkeit für den Zustand 1)

$$\frac{\exp\{x_{ij}\vartheta\}}{1+\exp\{\vartheta\}}$$

entsprechen, weswegen man auch von auto-logistischen *Modellen spricht. Abhängig von der Form, welche die lokalen Charakteristiken annehmen, spricht Besag [Bes74] dann auch von* auto-binomialen, auto-normalen (Auto-Gauß-) *oder* Auto-Poisson-*Modellen.*

2.5.2 Parametrisierte Potentiale

Wie bei den oben vorgestellten Auto-Modellen können Klassen von Zufallsfeldern dadurch beschrieben werden, daß man ihre Potentiale von Parametern abhängig macht. Es wird sich spätestens in den Abschnitten zur Parameterschätzung (3.1.3 und 4.5) als nützlich erweisen, diese Parametrisierung zu formalisieren.

Für eine Familie A_i, $i = 1, \ldots, n$, von Mengen bezeichne $\prod_{i=1}^{n} A_i$ das kartesische Produkt $A_1 \times \ldots \times A_n$ dieser Mengen.

Definition 2.18 (Parametrisiertes Potential)
Es seien L^* eine endliche Indexmenge,

$$\Gamma_p^* \subseteq \mathbb{R} \qquad \forall\, p \in L^* \tag{2.16}$$

und

$$\Gamma^* \subseteq \prod_{p \in L^*} \Gamma_p^*. \tag{2.17}$$

Dann heißt das Potential $V = V(\theta)$ durch $\theta \in \Gamma^*$ parametrisiert*, falls V durch θ eindeutig festgelegt ist. V heißt* endlich *bzw.* abzählbar parametrisiert*, falls Γ^* endlich bzw. abzählbar ist. L^* heißt* duales Gitter *und* Γ^* Parameterraum.

Jede *Parametrisierung* θ ist damit ein Vektor von reellen Zahlen, für den durch die Teilmengenbeziehungen (2.16) und (2.17) beliebige Einschränkungen gefordert werden können. Im Grunde reichte in (2.16) auch die Gleichheit, die Teilmengenbeziehung gestattet aber in der Regel eine übersichtlichere Formulierung der an den Parameterraum Γ^* geknüpften Bedingungen. Um die Abhängigkeit einer Gibbs-Verteilung mit Energiefunktion $U(\theta) = \sum_{C \in \mathcal{C}} V_C(\theta)$ von ihrer Parametrisierung auszudrücken, wird die Schreibweise

$$P_\theta(X = x) = \frac{1}{Z^\theta(T)}\, e^{-U(\theta,x)/T}$$

mit

$$Z^\theta(T) = \sum_{y \in \Gamma^L} e^{-U(\theta,y)/T}$$

vereinbart.

Wie das folgende Beispiel zeigt, genügt mit obiger Definition zur Festlegung einer Teilklasse von Gibbs-Feldern über einem festen Gitter und bezüglich eines festen Nachbarschaftssystems die Angabe eines parametrisierten Potentials V und des zugehörigen Parameterraums Γ^*.

Beispiel 2.19 (Auto-logistisch parametrisiertes Modell)
Das auto-logistische Modell aus Beispiel 2.17 ist parametrisiert durch die Indexmenge $L^* = L \cup L^2$ *und*

$$\Gamma^* = \left\{ \theta \in \prod_{p \in L^*} \Gamma_p^* \middle| \theta_{ij,kl} = \theta_{kl,ij} \right\}$$

mit

$$\Gamma_{ij}^* = I\!R \quad und \quad \Gamma_{ij,kl}^* = \begin{cases} I\!R & (k,l) \in \eta_{ij} \\ \{0\} & (k,l) \notin \eta_{ij}. \end{cases}$$

2.5.3 Kontrast-Modelle

Über die Auto-Modelle Besags hinaus wurden noch zahlreiche weitere, für die Bildverarbeitung interessante Teilklassen von Markoff-Feldern betrachtet [HS80, CrJ83, GG84, Der85, GMcC85, Cha88, SAG89, Gre90], von denen sich einige in einer neuen elegant verallgemeinern lassen.

In dieser neuen Teilklasse wird der Grauwertunterschied zweier benachbarter Bildpunkte als Kontrast (Intensitätsdifferenz) aufgefaßt und mittels einer *Kontrastfunktion* δ bewertet. Um mit der intuitiven Vorstellung konform zu gehen, muß die Klasse der zulässigen Bewertungsfunktionen etwas eingeschränkt werden:

Definition 2.20 (Kontrastfunktion)
Eine Funktion $\delta : \Gamma \times \Gamma \to [0,1]$ *heißt* Kontrastfunktion, *wenn sie für beliebige* $g, h, g', h' \in \Gamma$ *die folgenden Bedingungen erfüllt:*

(i)	$\delta(g,g) = 0$	*(Normiertheit)*
(ii)	$\lvert g-h \rvert \le \lvert g'-h' \rvert \Rightarrow \delta(g,h) \le \delta(g',h')$	*(Symmetrische Monotonie)*

Bedingung *(ii)* enthält insbesondere die für Cliquenpotentiale unerläßliche Symmetriebedingung. Kontrastfunktionen erinnern in ihrer Definition an Metriken, doch wird in *(ii)* anstelle der Dreiecksungleichung lediglich monotones Wachstum im Betrag der Differenz gefordert, um beispielsweise auch konvexe Bewertungen zuzulassen. Die neue Teilklasse zeichnet sich dadurch aus, daß aus ihren Cliquenpotentialen ablesbar ist, in welchem Maße ein mittels δ bewerteter Kontrast an der enstprechenden Bildstelle gefördert oder unterdrückt werden soll.

Definition 2.21 (δ-Kontrast-Zufallsfelder)
Es seien X *ein Markoff- bzw. Gibbs-Zufallsfeld über* L *mit Nachbarschaftssystem* η *und* δ *eine Kontrastfunktion.* $L^* = \Gamma \cup L^2$ *sei eine Indexmenge*

und

$$\Gamma^* = \left\{ \theta \in [-1,1]^{L^*} \,\Big|\, \sum_{g \in \Gamma} \theta_g = 0,\ \theta_{ij,kl} = \theta_{kl,ij} \quad \forall\, (i,j) \neq (k,l) \in L \right\}$$

ein Parameterraum. Hat dann X ein durch $\theta \in \Gamma^*$ parametrisiertes Potential $V(\theta)$, für das

$$V_C(\theta, x) = \begin{cases} \theta_{x_{ij}} & C = \{(i,j)\} \\ \theta_{ij,kl} \cdot \delta(x_{ij}, x_{kl}) & C = \{(i,j),(k,l)\},\ (i,j) \neq (k,l) \\ 0 & sonst \end{cases}$$

gilt, so ist X ein δ-Kontrast-Zufallsfeld (δ-CRF) *und V ein* (δ-)Kontrast-Potential.

Für die Komponenten der Parametrisierung θ eines δ-Kontrast-Zufallsfeldes wird die Bezeichnungskonvention

$$\theta = (\alpha_0, \ldots, \alpha_{G-1}, \beta_{(1,1),(1,1)}, \ldots, \beta_{(N,M),(N,M)})$$

vereinbart. Da eine Realisation eines Zufallsfeldes, also ein Bild, um so wahrscheinlicher wird, je geringer seine Energie ist, lassen sich δ-Kontrast-Modelle sehr anschaulich interpretieren. Die Parameter α_g geben an, wie sehr das Vorkommen eines bestimmten Grauwertes zur Wahrscheinlichkeit des Bildes beiträgt und können als externes Kraftfeld aufgefaßt werden. Je größer α_g ist, desto seltener wird der Grauwert g auftreten. Die *Bindungsstärke* $\beta_{ij,kl}$ bestimmt, ob ein durch δ bewerteter Kontrast zwischen den Grauwerten der Pixel (i,j) und (k,l) erwünscht ist ($\beta_{ij,kl} < 0$) oder nicht ($\beta_{ij,kl} > 0$). Durch die Form der Kontrastbewertung können z.B. harte (konvexes δ) oder weiche (konkaves δ) Grauwertübergänge gefördert werden.

Die jeweilige Normierung auf das Intervall $[-1,1]$ stellt wegen der noch freien Wahl der Temperatur T des Feldes keine Einschränkung dar[8].

Offensichtlich ist die Beschreibung eines δ-Kontrast-Feldes im allgemeinen aufwendiger als die bloße Angabe der Grauwertmatrix eines Bildes. Andererseits wird das Modell durch seinen stochastischen Charakter ohnehin erst sinnvoll, wenn größere Regionen der Bilder gleichartige Bindungsstärken aufweisen. Dann aber reduziert sich die Zahl der Parameter, wie das folgende Beispiel zeigt, ganz erheblich.

Beispiel 2.22 (Textur)
Eine Textur werde beschrieben durch ein homogenes Markoff-Feld. Unterstellt man als Nachbarschaftssystem η^2, so ergeben sich die in Abbildung 2.9

[8] vgl. Ausführungen nach Definition 2.11

gezeigten vier Bindungstypen $\beta_1, \ldots, \beta_4$. Für die Festlegung eines δ-Kontrast-Modells reicht es demnach aus, $\theta = (\alpha_0, \ldots, \alpha_{G-1}, \beta_1, \ldots, \beta_4)$ und δ anzugeben[9]*.*

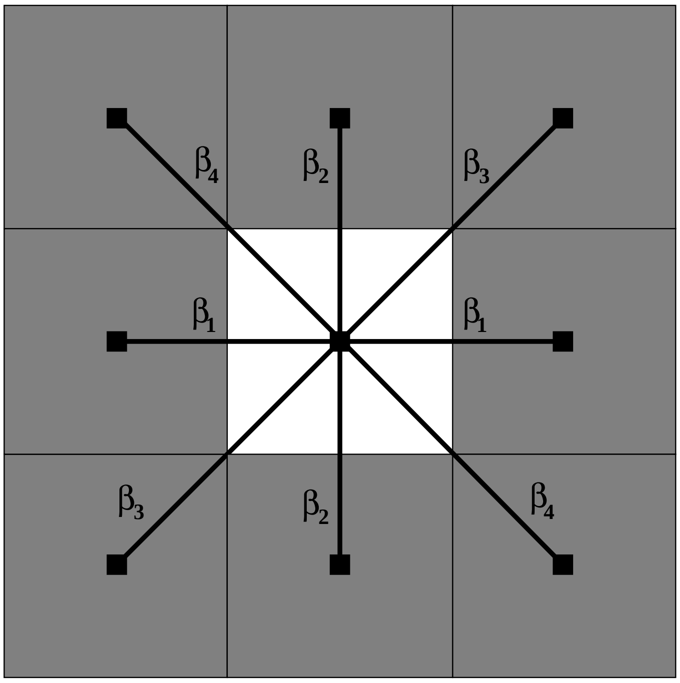


Abbildung 2.9: Bindungstypen in η^2

Wie im vorstehenden Beispiel sind Gibbs-Felder mit Kontrast-Potentialen offensichtlich genau dann homogen, wenn die jeweiligen Bindungsstärken zweielementiger Cliquen gleichen Typs identisch sind. Anhand einiger weiterer Beispiele soll verdeutlicht werden, daß die Klasse der δ-Kontrast-Modelle trotz ihrer einfachen Beschreibung eine echte Obermenge zahlreicher anderenorts verwendeter Klassen darstellt:

Beispiel 2.23 (Spezielle Kontrast-Modelle)

(i) Derin betont den potentiellen Nutzen einer Klasse von Markoff-Feldern, die in obiger Notation im wesentlichen durch eine Kontrastfunktion der Form

$$\delta(g,h) = \left(\frac{|g-h|}{G-1}\right)^n,$$

[9]Für eine genauere Erläuterung des Begriffs Textur und konkrete Beispiele wird auf Abschnitt 3.1 verwiesen.

$n \in I\!N$ ein weiterer Parameter, bestimmt ist. In [Der85] werden zahlreiche Beispiele für Realisierungen gegeben. Man beachte, daß sich die dort angegebene Form durch geeignete Wahl der Temperatur T in die obige überführen läßt.

(ii) Geman & Geman [GG84] benutzen zur Restauration gestörter Bilder eine besonders einfache Variante des Kontrast-Modells, indem sie $\alpha_g = 0$ für alle $g \in \Gamma$ wählen und für benachbarte $(i,j),(k,l)$ die Bindungsstärke $\beta_{ij,kl} = -\frac{2}{3}$ setzen. Als Kontrastfunktion wählen sie $\delta(g,h) = 1 - \delta_{g,h}$ ($\delta_{g,h}$ das Kronecker-Symbol). Beim Vergleich mit der von ihnen angebenen Form (vgl. Abschnitt 3.2.5) fällt auf, daß sich die Paarpotentiale jeweils um $\frac{1}{3}$ unterscheiden. Diese Differenz hat jedoch keinen Einfluß auf die Verteilung (2.9), da lediglich $e^{1/3}$ herauszukürzen ist.

(iii) Es läßt sich zeigen, daß die von Shridar et al. *in [SAG89] aufgeführte quadratische Potentialfunktion in ein Kontrast-Modell mit identischen $\beta_{ij,kl}$ für alle benachbarten Paare (i,j) und (k,l) sowie einer Kontrastfunktion*

$$\delta(g,h) = \left(\frac{g-h}{G-1}\right)^2$$

übertragbar ist.

(iv) Für die Rekonstruktion von tomographischen Schichtaufnahmen aus den gegebenen Projektionen werden a-priori-*Annahmen über die Verteilung der beobachteten Bilder von S. Geman und McClure [GMcC85] durch ein Kontrast-Modell modelliert, bei dem $\alpha_g = 0$ für alle $g \in \Gamma$, $\beta_{ij,kl} = \beta$ für direkt benachbarte Pixel $(i,j),(k,l)$ (also solche mit gemeinsamer Kante) und $\beta_{ij,kl} = \beta\sqrt{1/2}$ für zwei benachbarte Pixel mit lediglich gemeinsamem Eckpunkt sind. δ hat die Form*

$$\delta(g,h) = \left(1 + \left(\frac{\varrho}{g-h}\right)^2\right)^{-1}.$$

Dabei sind β und ϱ die verbleibenden Parameter. Je nach deren Wahl muß die Normierung von Kontrastfunktion und Kontrast-Parametern gegebenenfalls durch Anpassung der Temperatur T erfolgen. In [Gre90] wandelt Green das Modell dahingehend ab, daß er die Kontrastfunktion durch

$$\delta(g,h) = c_1 \log\cosh\left(c_2 \cdot \frac{g-h}{\varrho}\right)$$

mit Parametern c_1, c_2 und ϱ ersetzt, weil die Energiefunktion dadurch eine für die nötigen Berechnungen günstigere Form annimmt (vgl. Abschnitt 3.4).

Es sind jedoch auch Modelle verwendet worden, die nicht zu Kontrast-Modellen umformuliert werden können. Das in Beispiel 2.17 erwähnte auto-binomiale Modell wird von Cross und Jain [CrJ83] für Nachbarschaften verschiedener Ordnung parametrisiert und zur Charakterisierung von Texturen verwendet. In Abschnitt 3.1 wird ihre Modellierung vorgestellt und man erkennt, daß es im allgemeinen keine *direkte* Entsprechung unter den Kontrast-Modellen gibt, weil in den Cliquenpotentialen Produkte von Grauwerten auftreten müßten, was wegen Bedingung *(ii)* in Definition 2.20 nicht geht. Ähnliches gilt für die übrigen in Beispiel 2.17 genannten Auto-Modelle, insbesondere also für die wichtigen auto-normalen.

Daß die Klasse der Kontrast-Zufallsfelder dennoch zur Beschreibung einer beachtlichen, möglicherweise sogar genügenden Anzahl von Bildtypen ausreicht, wird in Abschnitt 3.1.2 zu untermauern versucht, indem Beispiele und Hinweise zur Modellierung spezifischer Bildinhalte gegeben werden.

Kapitel 3

Charakteristische Anwendungen in der Bildverarbeitung

Lustige Person: In bunten Bildern wenig Klarheit,
viel Irrtum und ein Fünkchen Wahrheit.
J. W. von Goethe: Faust,
Vorspiel auf dem Theater

In diesem Kapitel soll an beispielhaften Anwendungen gezeigt werden, daß Markoff-Felder ein leistungsfähiges Konzept zur Modellierung und (meist näherungsweisen) Lösung von zahlreichen Problemen der Bildverarbeitung sind. Die Anwendungen wurden daher nicht nur aus verschiedenen Gebieten ausgewählt, sondern auch so, daß sie unterschiedliche Ausprägungen und Möglichkeiten der Verwendung von Markoff-Feldern erkennen lassen.

Während Texturen recht allgemeine Behandlung erfahren, werden die konkreteren Probleme Textursegmentierung, Bildrestauration und Bildrekonstruktion jeweils anhand eines ausgewählten Lösungsvorschlages dargestellt. In den entsprechenden Abschnitten sollten auch die von den verschiedenen Autoren vorgenommenen Modifikationen, Einschränkungen oder Betonungen des Bildmodells deutlich werden. Jeder dieser Darstellungen sind abschließend einige Anmerkungen beigefügt.

Dieses Kapitel gibt damit nicht nur einen Überblick über diverse Einsatzgebiete von Markoff-Feldern, sondern auch über Lösungstechniken von allgemeinerem Interesse.

3.1 Textur

Bilder und die darin enthaltenen Objekte bestehen in den seltensten Fällen aus großen zusammenhängenden Flächen gleichen Grauwertes. Ein wichtiger Aspekt der Bildanalyse ist daher ein ausreichendes Verständnis des Textur-Konzepts [Acu92], denn in Texturen sind häufig nützliche Informationen über Form, Lage und Anordnung von Bildobjekten enthalten.

Die vier herausragenden Aufgabenstellungen, bei deren Lösung Texturen eine tragende Rolle spielen, sind Klassifikation, Segmentierung, Computergrafik und Bildcodierung [CrJ83, Querverweise darin]. Aufgrund der zentralen Bedeutung dieser und verwandter Anwendungen sind Texturen ein erster Prüfstein für die Adäquatheit des Markoff-Modells.

Vorab ist jedoch die Frage zu klären:

3.1.1 Was ist Textur?

Leider kann hier keine erschöpfende Antwort gegeben werden, denn Textur ist in erster Linie ein intuitiver Begriff. Er bezeichnet Bilder oder Bildbereiche mit einem einheitlichen Erscheinungsbild, einer durchgängigen Struktur aus mehr oder weniger angeordneten Grundelementen. Der ganzheitliche Eindruck entsteht dadurch, daß keines dieser Grundelemente dem Betrachter in besonderer Weise auffällt [GDO85] (Abbildung 3.1).

404

Ausgehend von dieser vagen Vorstellung kann man in umgekehrter Vorgehensweise ein Modell vorgeben und alles, was dieser Modellvorstellung genügt, Textur nennen. Texturen lassen sich dann zunächst durch die Zugehörigkeit zu einer von zwei wesentlichen Modellkategorien unterscheiden [GDO85, Wei93]:

- *Deterministische Texturen:* Elementare Bildeinheiten, sogenannte Primitive, werden gemäß einer Anordnungsvorschrift plaziert (z.B. Kinderzimmertapete, Schachbrett).
- *Stochastische Texturen:* Die charakteristischen Eigenschaften sind nur durch statistische Kenngrößen (wie jene aus Abschnitt 1.2) beschrieben (z.B. Schnee auf dem Fernsehbildschirm, Wiese).

Darüber hinaus sind natürlich auch Mischformen möglich, etwa durch stochastische Anordnungsvorschriften oder Primitive. Weiterhin wäre eine Hierarchie von Texturen denkbar.

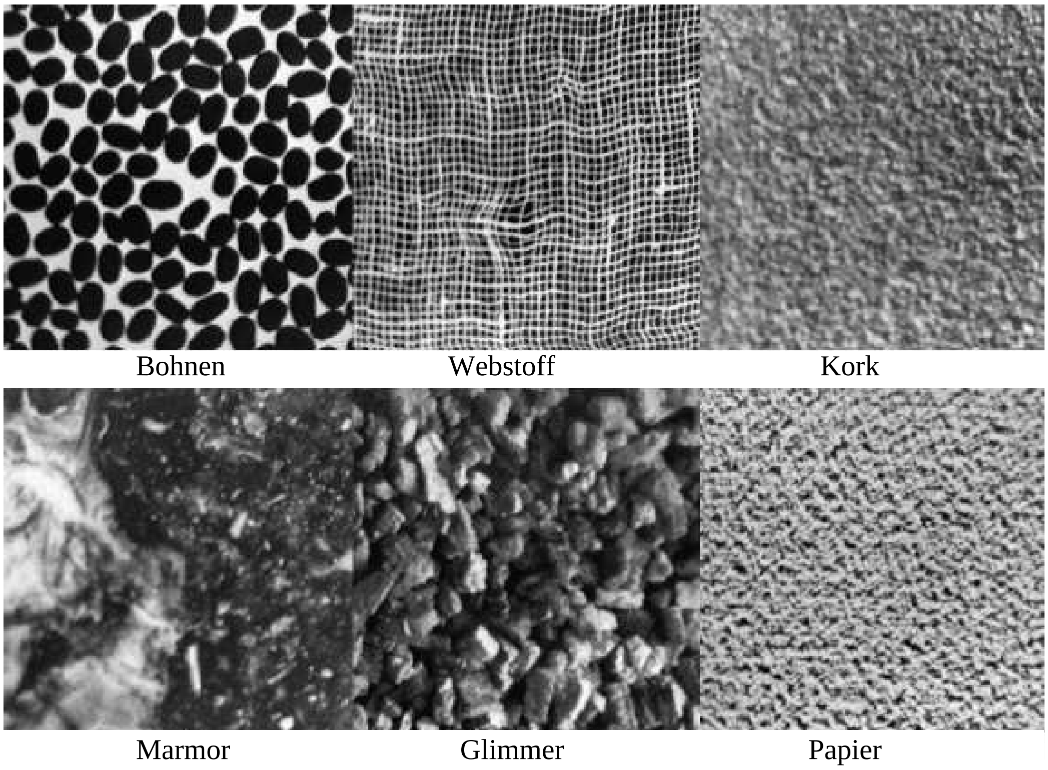


Abbildung 3.1: Texturbeispiele aus [Bro66]

Für diese Arbeit sind naheliegenderweise nur Texturen von Interesse, die nicht rein deterministisch sind. Zum Zwecke ihrer Beschreibung existieren eine Vielzahl von Modellen [GDO85, Querverweise darin][HS80, CrJ83] und eine ebensogroße Zahl von Merkmalen [Har79, Jäh91], welche meist aus Statistiken zweiter Ordnung (Abschnitt 1.2) abgeleitet sind, weil diese das menschliche Sehen bestimmen [Jul75].

Die hier getroffene Modellannahme folgt der von Hassner und Sklansky in [HS80] vorgeschlagenen und besteht darin, daß eine Textur ausschließlich durch ihre lokalen stochastischen Eigenschaften bestimmt ist. Realisierungen von Texturen entstehen dann durch Ziehungen aus entsprechenden Wahrscheinlichkeitsverteilungen.

Definition 3.1 (Textur)
Ein homogenes Markoff-Feld heißt Textur.

Dieses Postulat scheint überraschend einfach und möglicherweise zu restriktiv zu sein. Um die Relevanz des Modells zu untermauern, werden deshalb im nächsten Abschnitt allgemeine Konstruktionsprinzipien für Texturen vorgestellt. Der darauffolgende Abschnitt ist der Rückgewinnung der Parameter eines unterstellten Markoff-Modells aus einer beobachteten Textur gewidmet.

3.1.2 Synthese von Texturen

Eine elegante Methode zur Gewinnung realistischer Computergrafiken besteht darin, ein Objekt nicht als farbige Punktmenge, sondern lediglich durch seine Form und Position sowie Texturen, mit denen seine Oberflächen überzogen werden, zu beschreiben.

Da Texturen hier durch Markoff-Felder modelliert werden, stellt sich die Frage, wie das gewünschte Erscheinungsbild einer solchen Objektoberfläche in einem Potential codiert werden kann. Sicher ist es nicht sinnvoll, das Markoff-Feld durch explizite Angabe aller lokalen Charakteristiken (2.3) oder Cliquenpotentiale (2.5) festzulegen.

Für eine kompakte Repräsentation bieten sich parametrisierte Potentiale an. In Abschnitt 2.5 ist mit dem Kontrast-Potential ein solches vorgestellt worden, doch bevor auf die Eigenschaften der Kontrast-Modelle näher eingegangen wird, lohnt der Seitenblick auf ein bekanntes und gut untersuchtes Modell:

Das autobinomiale Modell

Da Hassner und Sklansky [HS80] ihre Untersuchungen auf binäre Markoff-Felder mit kleinen Nachbarschaften beschränkten, die durch wenige Parameter charakterisiert sind, wird stattdessen die von Cross und Jain [CrJ83] benutzte Charakterisierung näher betrachtet. Sie parametrisieren Texturen durch das von Besag [Bes74] vorgestellte autobinomiale Modell (siehe auch Beispiel 2.17) und verwenden dazu das hierarchische Nachbarschaftssystem η^n, $n = 1, \ldots, 4$, aus Beispiel 2.5 *(v)*[1]. Die binomialverteilten lokalen Charakteristiken haben die Form

$$P(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}) = \binom{G-1}{x_{ij}} \tau^{x_{ij}} (1-\tau)^{G-1-x_{ij}},$$

wobei τ die von der Nachbarschaft bestimmte "Erfolgswahrscheinlichkeit" für die Überwindung einer Grauwertstufe bei $G-1$ Versuchen ist. Zur Berechnung von τ wird zunächst für das jeweilige Nachbarschaftssystem η^n der implizite Parameter $\gamma(n)$ bestimmt:

$$\begin{array}{rclclcl}
\gamma(1) & = & \theta_{0,0} & + & \theta_{1,1}(x_{i-1,j} + x_{i+1,j}) & + & \theta_{1,2}(x_{i,j-1} + x_{i,j+1}) \\
\gamma(2) & = & \gamma(1) & + & \theta_{2,1}(x_{i-1,j+1} + x_{i+1,j-1}) & + & \theta_{2,2}(x_{i-1,j-1} + x_{i+1,j+1}) \\
\gamma(3) & = & \gamma(2) & + & \theta_{3,1}(x_{i-2,j} + x_{i+2,j}) & + & \theta_{3,2}(x_{i,j-2} + x_{i,j+2}) \\
\gamma(4) & = & \gamma(3) & + & \multicolumn{3}{l}{\theta_{4,1}(x_{i-1,j+2} + x_{i-2,j+1} + x_{i+1,j-2} + x_{i+2,j-1})} \\
 & & & + & \multicolumn{3}{l}{\theta_{4,2}(x_{i-1,j-2} + x_{i-2,j-1} + x_{i+1,j+2} + x_{i+2,j+1})}
\end{array}$$

[1] Nachbarschaftssysteme höherer Ordnung wurden von Cross und Jain nicht behandelt, weil deren Parameter aus den verwendeten 64x64 Pixel großen Bildern nicht mehr geschätzt werden können.

Wegen der speziellen Form der Parameter werden die Nachbarschaften η_{ij}^n so modifiziert, als sei das Gitter L ein Torus. Dies ist eine wenig natürliche, aber trotzdem weitverbreitete Annahme, die der aus der Fourieranalyse bekannten periodischen Fortsetzung des Bildes in alle Richtungen entspricht. Mit Hilfe von $\gamma(n)$ wird nun der Parameter τ für das Nachbarschaftssystem η^n, $n = 1, \ldots, 4$ durch

$$\tau = \frac{e^{\gamma(n)}}{1 + e^{\gamma(n)}}$$

definiert.

Die eigentlichen Parameter einer Textur sind demnach die Gewichte θ_{rs}. Positive Werte für ein θ_{rs} bewirken, daß der Grauwert x_{ij} zur Höhe derjenigen Grauwerte tendiert, deren Summe den Faktor hinter θ_{rs} ausmacht ("Anziehung", *attraction*). Negative Werte bewirken das Gegenteil ("Abstoßung", *repulsion*). Insgesamt kontrollieren die Gewichte also die Bildung von Gebieten gleichen Grauwerts (*clustering*). Weil die aus Summen bestehenden Faktoren in den $\gamma(n)$ jeweils eine Richtung vorgeben, in der das zugehörige Gewicht wirkt, spricht man von einer *isotropen* Textur, wenn $\theta_{r,1} = \theta_{r,2}$ für alle r gilt. Andernfalls nennt man sie *anisotrop*.

Cross und Jain geben einige Beispiele an, an denen die Wirkung der Parameterwahl ablesbar ist [CrJ83]. An dieser Stelle jedoch wird darauf verzichtet und auf die Darstellung des Kontrast-Modells im nächsten Unterabschnitt verwiesen.

Für die Generierung ihrer Beispiele benutzen Cross und Jain eine Variante des Simulated Annealing, die jeweils zwei Bildpunkte auswählt und unter bestimmten Bedingungen deren Grauwerte vertauscht. Auf die Einzelheiten wird hier nicht eingegangen, weil Berechnungsfragen in Kapitel 4 ausführlicher behandelt werden. Von Bedeutung ist aber, daß die relativen Häufigkeiten der Grauwerte bei dieser Methode erhalten bleiben, wodurch das Histogramm des Ergebnisbildes schon bei der Erzeugung des Startbildes festgelegt wird.

Um die Ausbildung des Histogramms flexibel zu halten, hat Acuna [Acu92] das autobinomiale Modell der obigen Form dahingehend abgewandelt, daß die lokalen Charakteristiken die Gestalt

$$P(X_{ij} = x_{ij} | X_{L\setminus\{(i,j)\}} = x_{L\setminus\{(i,j)\}}) = \frac{\binom{G-1}{x_{ij}} \tau^{x_{ij}} (1-\tau)^{G-1-x_{ij}} + \sigma \cdot H_{ij}(x,h)}{Z_{ij}}$$

annehmen. Hierbei ist $h : \Gamma \to I\!N$ ein Histogramm und

$$H_{ij}(x,h) := 2h(x_{ij}) - \frac{2}{N \cdot M} \sum_{(k,l) \neq (i,j)} \delta_{x_{ij}, x_{kl}}$$

mit dem Kroneckersymbol $\delta_{x_{ij},x_{kl}}$. Der zweite Summand im Zähler mißt die Abweichung vom vorgebenen Histogramm und bewertet durch den Faktor $\sigma > 0$, wie verbindlich die Vorgabe h ist.

Obwohl dadurch die Nachbarschaften auf das ganze Gitter L ausgedehnt werden (Beispiel 2.5 *(ii)*), bleiben die auftretenden Berechnungen lokal und effektiv durchführbar. Der von Acuna zur Texturgenerierung verwendete Metropolis-Algorithmus [MRT53] ist in Abschnitt 4.2.1 formalisiert. Es wird jeweils nur ein Bildpunkt (i, j) ausgewählt und für diesen ein neuer Grauwert vorgeschlagen. Der Algorithmus benötigt die für eine Berechnung zu aufwendigen normierenden Konstanten Z_{ij} nicht, und die $H_{ij}(x, h)$ können leicht aus einer globalen Zählung, die anschließend jeweils entsprechend den neuen Grauwerten zu aktualisieren ist, hergleitet werden.

Ein grundlegender Nachteil der Binomialverteilung als lokaler Charakteristik ist der etwas unscharfe Eindruck, den die generierten Texturrealisierungen hervorrufen. Fehlen nämlich Parameter höherer Ordnung, um sanfte Übergänge zwischen sehr verschiedenen Grauwerten zu verhindern, resultiert die Unimodalität der Binomialverteilung in einer Art Tiefpaßfilter-Effekt [CrJ83]. Desweiteren sind die Parameter in ihrer relativen Größe wenig aussagekräftig, von der Temperatur wird nicht einmal Gebrauch gemacht. Demgegenüber bieten Kontrastmodelle eine nachvollziehbare Parametrisierung, zu welcher der Temperaturparameter sogar eine offensichtliche Klassenbildung bewirkt.

Das Kontrast-Modell

Durch einfache Beispiele sollen die Möglichkeiten der Texturgenerierung mit Hilfe der in Abschnitt 2.5 eingeführten Kontrast-Modelle vorgeführt werden. Aus den speziellen Bildeigenschaften, die hier isoliert betrachtet werden und als Basisbausteine dienen können, sind erheblich komplexere Texturen durch Kombination oder Überlagerung zu erzeugen.

Die in den Abbildungen 3.2–3.7 gezeigten Beispiele wurden mit dem in Anhang A beschriebenen Programm `gm` generiert. Um das Verhältnis der Grauwertvorkommen zueinander ausgewogen zu halten, wurde in allen Beispielen $\alpha_g = 0$, $g \in \Gamma$, gesetzt. Zum besseren Verständnis der Parametrisierungen rufe man sich die nach Definition 2.21 gegebenen Interpretationen noch einmal ins Gedächtnis zurück.

Zunächst wird anhand von binären Texturen der Einfluß der Bindungsstärken $\beta_{ij,kl}$ dargestellt. Da für $\Gamma = \{0, 1\}$ jede Kontrastfunktion von der Form

$$\delta(g, h) = \gamma \cdot \delta_{g,h}$$

für ein $0 \leq \gamma \leq 1$ ist, $\gamma = 0$ aber auf unabhängige Komponenten des Zufallsfeldes führt, wird o.E. $\delta(g,h) = \delta_{g,h}$ gewählt.

Das erste Beispiel dient vornehmlich dazu, die Bedeutung des Temperaturparameters zu unterstreichen:

- *Anziehung* (*attraction/clustering*, Abb. 3.2): Das verwendete Nachbarschaftssystem ist η^2 mit Bindungsstärken $\beta_1 = \ldots = \beta_4 = 1$ (vgl. Abbildung 2.9). Die sich daraus ergebende Energiefunktion fördert die Ausbildung homogener Regionen. In der Abbildung ist deutlich zu erkennen, wie der Temperaturparameter T den Einfluß der Energie auf die Verteilung einer Textur verändert (vgl. Ausführungen nach Definition 2.11).

Um die anhand der nachfolgenden Gestaltungsmerkmale dargestellten Modelleigenschaften nicht durch hohe Temperaturen zu verwischen, wurde bei allen weiteren Realisierungen jeweils $T = 0.01$ gewählt. Die verwendeten Parameter sind graphisch so dargestellt, daß jeder Nachbar (k,l) des Punktes (i,j) mit der Bindungsstärke $\beta_{ij,kl}$ beschriftet ist. Insbesondere läßt sich aus den aufgeführten Nachbarn das Nachbarschaftssystem ablesen:

- *Abstoßung* (*repulsion*, Abb. 3.3): Die negativen Werte für horizontale und vertikale Bindungen in Bild a) unterdrücken die Bildung gleichgefärbter Pixelpaare in diesen Richtungen. Es bilden sich daher schachbrettartig gemusterte Regionen mit instabilen Grenzen aus. In b) sind die Bindungsstärken mit allen Nachbarn negativ, so daß sich in der Pixelebene parallele und rechtwinklige Teilmuster ergeben, wohingegen die Wahl der Parameter in c) bewirkt, daß es mehrere "stabile" Musterungen gibt. Diese sind durch die fehlenden Bindungen zu den unmittelbaren Nachbarn gröber als bei a) und b).

- *Verwerfung* (*attraction-repulsion*, Abb. 3.4): Durch Bindungsstärken mit unterschiedlichem Vorzeichen bei unterschiedlich nahen Nachbarn wird die lokale Bildung von Blöcken mit wachsender Größe sofort wieder unterdrückt.

- *Gerichtetheit* (*anisotropy*, Abb. 3.5): Wird die Gleichfarbigkeit von Punkten nur in einer Richtung gefördert, so erhält man anisotrope Texturen. Vergleicht man a) und b), so erkennt an, daß diese umso gröber werden, je größer die Summe der beteiligten Bindungsstärken ist. Wird dagegen Kontrast in der zur Ausrichtung senkrechten Richtung unterstützt, bilden sich wie in c) feinere Strukturen mit dünneren Linien aus.

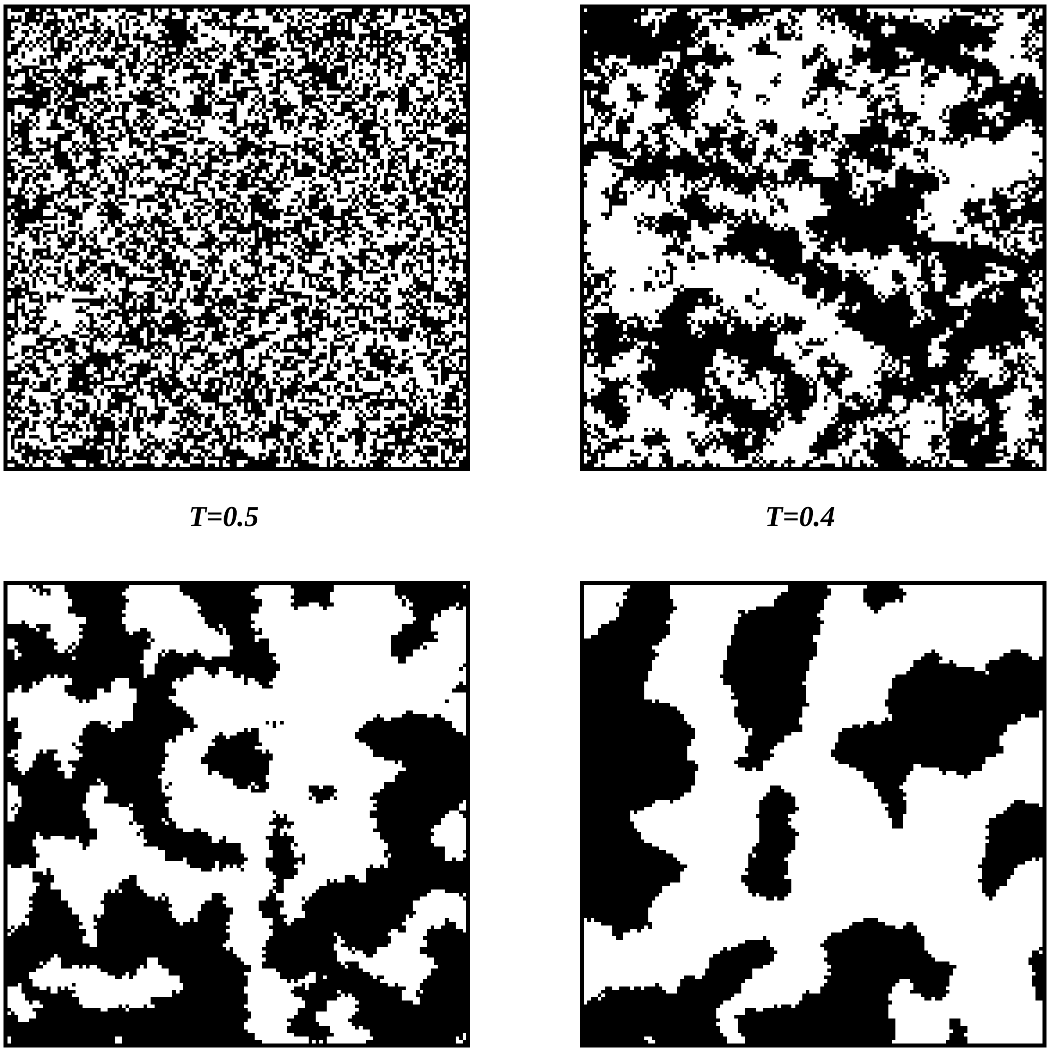


Abbildung 3.2: Auswirkung des Temperaturparameters

Positive Bindungsstärken in alle Richtungen bewirken Clusterung, die mit abnehmendem Temperaturparameter immer ausgeprägter wird.

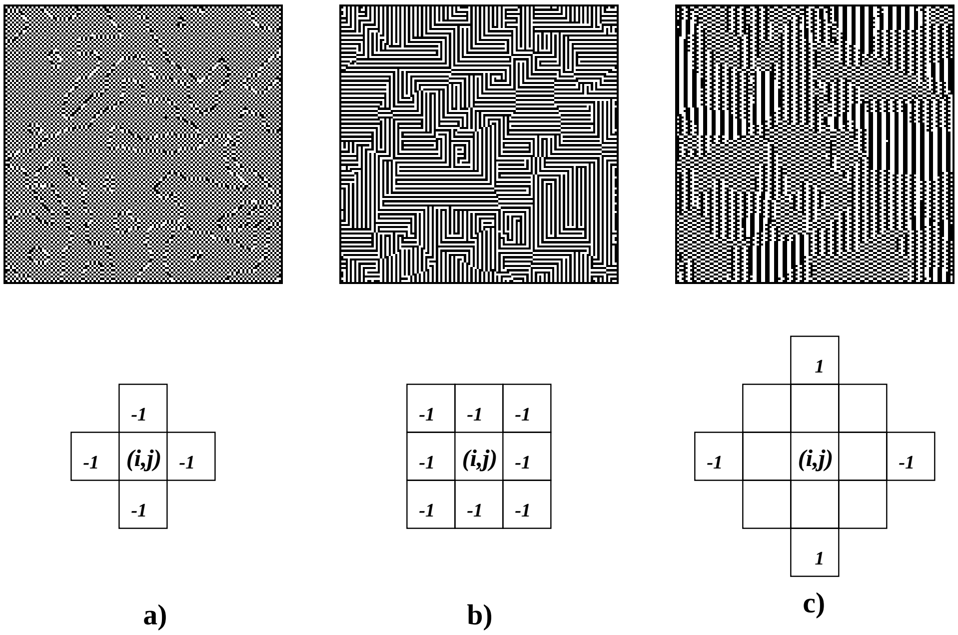


Abbildung 3.3: Abstoßungsmuster durch negative Bindungsstärken

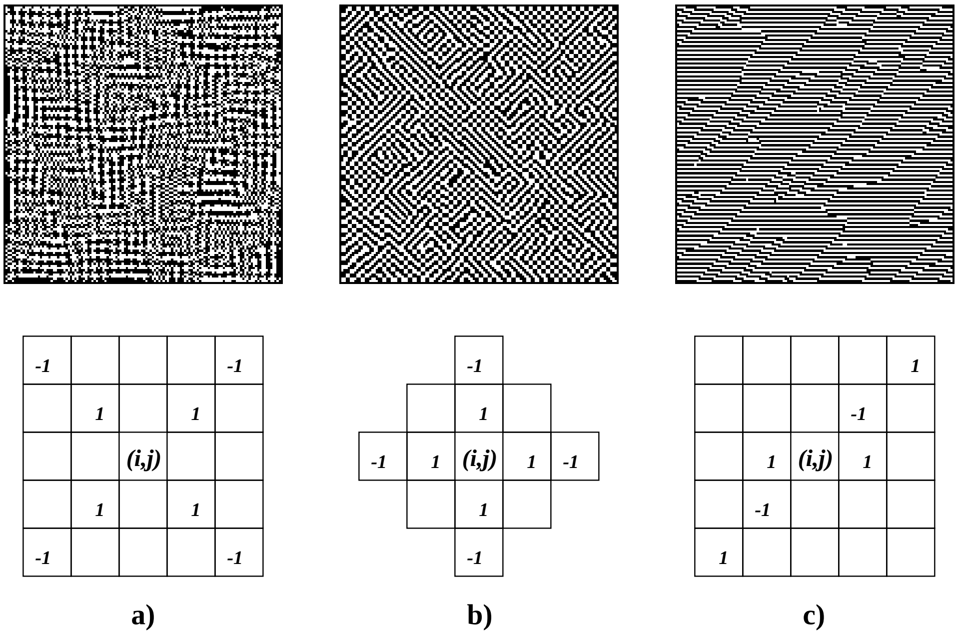


Abbildung 3.4: Verwerfungen durch gegensätzliche Bindungsstärken

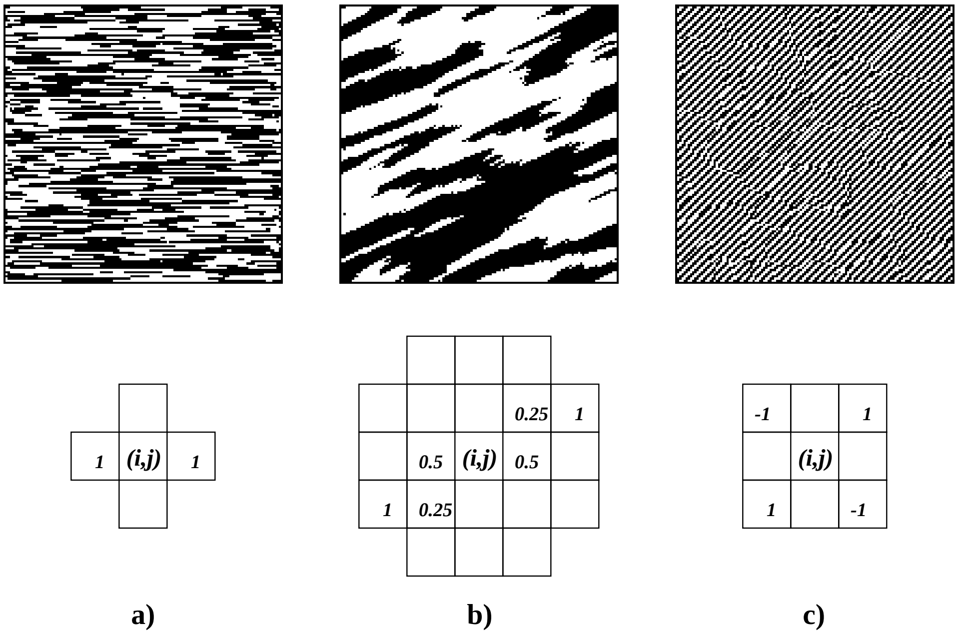


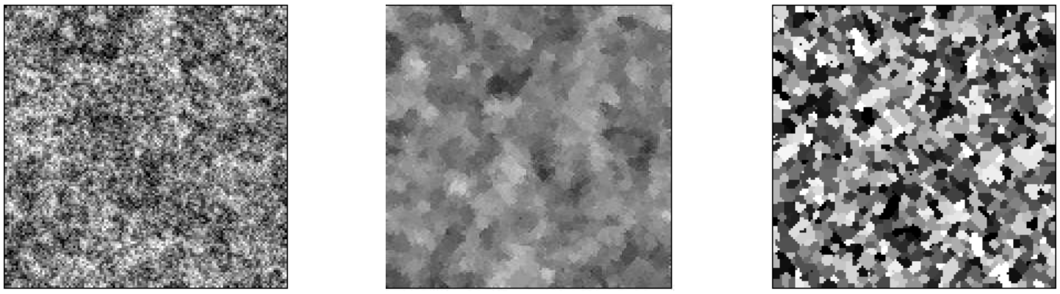
Abbildung 3.5: Anisotrope Texturen unterschiedlicher Ausrichtung

Für Bilder mit mehr als zwei Grauwerten zeigt Abbildung 3.6 die Auswirkungen verschiedener Kontrastfunktionen bei gleicher Parametrisierung $\theta = (0, \ldots, 0, 1, 1, 1, 1)$. Die Realisierung des $\left(\frac{|g-h|}{G-1}\right)^2$-Kontrast-Modells in a) weist zwar wie in Abbildung 3.2 eine Clusterung auf, durch die konvexe Form der Kontrastfunktion sind jedoch auch sprunghaftere Übergänge möglich. In b) dagegen bewirkt die konkave Kontrastfunktion deutlich weichere Übergänge, die in c) auf die Spitze getrieben werden. Da dort jeder von 0 verschiedene Grauwertunterschied mit 1 bewertet wird, bilden sich zusammenhängende Regionen gleichen Grauwertes, von denen beliebige aneinander grenzen können.

Abbildung 3.7 zeigt zwei Beispiele für Texturrealisierungen mit mehreren Grauwerten und einigen der oben angesprochenen Eigenschaften. Die Kontrastfunktion ist hier $\delta(g, h) = \frac{|g-h|}{G-1}$.

3.1.3 Analyse von Texturen

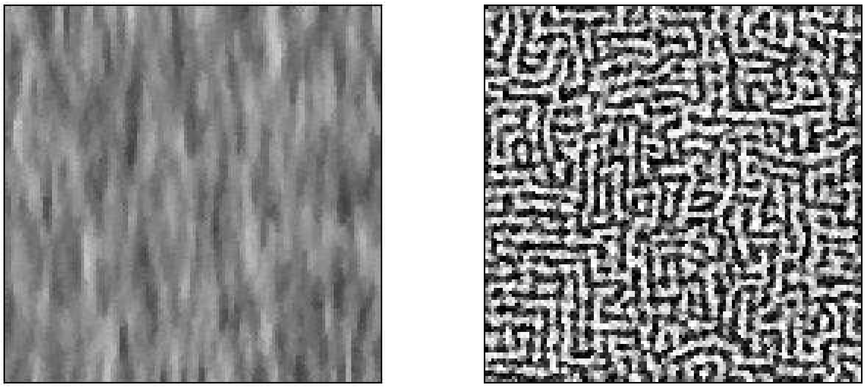
Es ist aus zwei Gründen interessant, aus einem gegebenen Bild die Parameter eines unterstellten Modells zu schätzen. Zum einen ergeben sich daraus Hinweise auf geeignete Parametersätze zur Generierung beabsichtigter Texturen, und zum anderen ist die Kenntnis des unterliegenden Modells für fast alle der in den folgenden Abschnitten angeschnittenen Problemlösungen von essentieller Bedeutung.



a) b) c)

Abbildung 3.6: Auswirkungen der Kontrastfunktion

Clusterung bei 32 Grauwerten mit unterschiedlichen Kontrastfunktionen: a) $\delta(g,h) = \left(\frac{|g-h|}{G-1}\right)^2$, b) $\delta(g,h) = \sqrt{\frac{|g-h|}{G-1}}$, c) $\delta(g,h) = \delta_{g,h}$.



		1		
	0.5	*1*	*0.5*	
-0.5	*0.5*	*(i,j)*	*0.5*	*-0.5*
	0.5	*1*	*0.5*	
		1		

-0.3	*-0.3*	*-0.3*	*-0.3*	*-0.3*
-0.3	*0.75*	*1*	*0.75*	*-0.3*
-0.3	*1*	*(i,j)*	*1*	*-0.3*
-0.3	*0.75*	*1*	*0.75*	*-0.3*
-0.3	*-0.3*	*-0.3*	*-0.3*	*-0.3*

Abbildung 3.7: Komplexere Texturrealisierungen mit 32 Grauwerten

Leider ist das Problem der Parameterschätzung jedoch noch nicht befriedigend gelöst. Die nächsten Unterabschnitte stellen zwei der gängigen Verfahren vor und weisen auf die inhärenten Probleme hin.

Codings

Wie vieles andere im Zusammenhang mit Zufallsfeldern auch geht die *coding method* auf die große Bestandsaufnahme von Besag [Bes74] zurück. Für ein gegebenes Bild x und ein bekanntes oder vermutetes Bildmodell, bestehend aus Nachbarschaftssystem und parametrisiertem Potential, wird eine Maximum Likelihood-Schätzung (ML-Schätzung) der Parameter gesucht. Die dabei erhaltene Parametrisierung maximiert für das gegebene Bild die Wahrscheinlichkeit einer Teilmenge der Komponenten, bedingt unter dem Rest des Zufallsfeldes.

Weil die Komponenten des Zufallsfeldes nicht unabhängig sind, besteht der erste Schritt aus der Partitionierung des Gitters in möglichst große Teilgitter, die sogenannten *codings*. Sie ist abhängig vom Nachbarschaftssystem und erfolgt derart, daß kein Nachbar eines Bildpunktes auf demselben Teilgitter liegt wie dieser Punkt. Abbildung 3.8 zeigt entsprechende Partitionierungen für die Nachbarschaftssysteme η^1 und η^2, wobei Elemente desselben *codings* gleich eingefärbt sind.

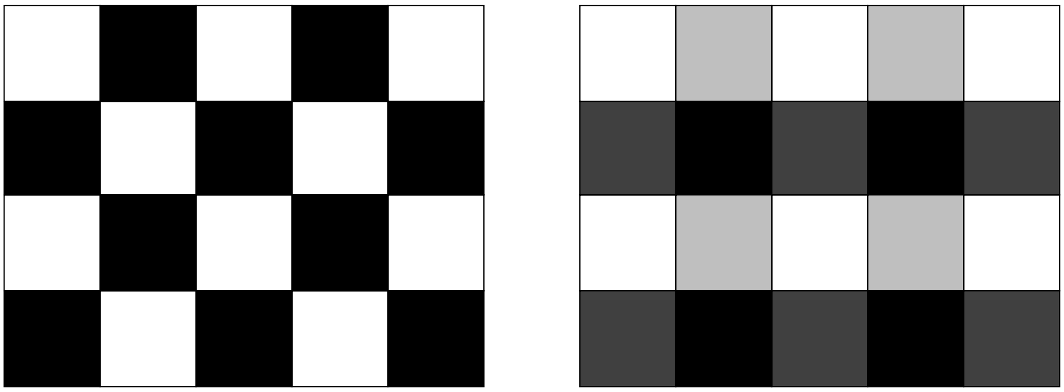


Abbildung 3.8: *Codings* für η^1 und η^2

Für ein *coding* $\zeta \subseteq L$ besteht die ML-Schätzung aus derjenigen Parametrisierung θ, für die

$$P_\theta(X_\zeta = x_\zeta | X_{L\setminus\zeta} = x_{L\setminus\zeta})$$

maximal wird. Da die Elemente eines *codings* unabhängig sind, kann obige Wahrscheinlichkeit faktorisiert werden, woraus sich für die log-Likelihood

$$\sum_{(i,j)\in\zeta} \log P_\theta(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}) \tag{3.1}$$

ergibt. Die Logarithmierung stellt keine Komplikation, sondern wegen der speziellen Form der Gibbs-Verteilung (2.11) meist eine Vereinfachung der Berechnung dar.

Leider hat die *coding method* zwei wesentliche Nachteile:

- *Nichtlinearität:* Die log-Likelihood (3.1) repräsentiert ein System nichtlinearer Gleichungen, deren Lösung wegen der meist komplizierten Zustandssummen im allgemeinen sehr aufwendig ist.

- *Ineffizienz:* Es werden nur Teile der Daten für die Bestimmung einer Schätzung verwendet. Mit größer werdenden Nachbarschaften werden die Abhängikeiten immer komplexer, die *codings* jedoch immer kleiner. Darüber hinaus gibt es keinen theoretisch gerechtfertigten Weg, die zu den verschiedenen *codings* erhaltenen Schätzungen zu kombinieren (meist wird einfach gemittelt).

Dessen ungeachtet ist die *coding method* ein gängiges Verfahren und wurde auch in [CrJ83] für die Texturanalyse benutzt. Offensichtlich beeinflußt bei festgehaltener Energiefunktion die Wahl des Nachbarschaftssystems das Ergebnis der Schätzung. Cross und Jain führen daher für verschiedene Nachbarschaftssysteme einen auf der χ^2-Methode (z.B. [Kre90]) basierenden *goodness of fit*-Test durch. Allerdings wies schon Besag [Bes74] darauf hin, daß die Schätzungen nur dann vergleichbar seien, wenn sie mit den gleichen Partitionierungen durchgeführt werden, was die Effizienz für das System mit den kleineren Nachbarschaften weiter verringert.

Maximum Pseudo-Likelihood

Derin und Elliott schlugen ein anderes Schätzverfahren vor, das lediglich Standardverfahren zur Lösung linearer Gleichungen benötigt [DE87]. Es entbehrt jedoch einer theoretisch fundierten Interpretation und ist für größere Nachbarschaften nicht mehr anwendbar, weil dann im allgemeinen zu wenige Beispiele der möglichen Nachbarschaftskonfigurationen auftreten.

Zusammen mit Lakshmanan verwendete Derin daher später eine andere, ebenfalls von Besag vorgeschlagene Methode [Bes86, LD89]. Anstelle der Partitionierung des Gitters in unabhängige *codings* wird das Produkt aller lokalen Charakterisitiken, die *Pseudo-Likelihood-Funktion*

$$\prod_{(i,j)\in L} P_\theta(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}}), \tag{3.2}$$

bezüglich der Parameter maximiert. Offensichtlich ist die Maximum Pseudo-Likelihood (MPL) Schätzung in der Regel auch nur eine Näherung. Lakshmanan und Derin berichten aber, ebenso wie Chalmond [Cha88], von ermutigenden Resultaten.

Die Maximierung von (3.2) kann mitunter erhebliche Berechnungsprobleme aufwerfen. Für die (angenäherte) Maximierung mittels Simulated Annealing wird in Abschnitt 4.5 eine abstrakte Beschreibung durch ein Maschinenmodell angegeben.

3.1.4 Anmerkungen

Für eine weitere Verbesserung der Texturbeschreibung durch Markoff-Felder scheint der in [ChJ93, Kapitel 6] dargestellte *multiresolution*-Ansatz besonders interessant zu sein. Anstelle *eines* Markoff-Feldes zur Beschreibung der gesamten Textur wird eine *Hierarchie* von Markoff-Feldern verwendet. Diese Hierarchie ist nach Auflösung (Gittergröße) gestaffelt, wodurch nicht nur ein effizienteres Generieren, sondern bei komplexen Texturen auch eine bessere Modellierung möglich zu sein scheint.

Neuere Anwendungen von durch Markoff-Felder modellierten Texturen in der Bildverarbeitung schließen 3D-Grafiken und 3D-Formerkennung ein. Einige davon sind in den entsprechenden Kapiteln von [ChJ93] behandelt.

Die üblichen Anwendungen in Medizin, Erderkundung oder Robotersteuerung erfordern meist die Erkennung oder Trennung von Objekten anhand von Form oder Lage. Daneben kann Texturanalyse aber auch zur automatischen Detektion von Produktionsfehlern in Web- und Faserstoffen dienen, weil die ansonsten regelmäßige Textur an den Fehlerstellen unterbrochen ist.

3.2 Bildrestauration

Eine der wichtigsten Anwendungen in der Bildverarbeitung ist der Versuch, ein gestörtes Bild wieder näher an den Originalzustand heranzubringen, denn der Abbildungsprozeß einer Szene auf ein Bild ist prinzipiell fehlerbehaftet. Selbst wenn die Restauration nicht das eigentliche Ziel der Anwendung ist, so ist sie häufig auch dann noch als Vorverarbeitungsschritt nötig, um eine weitergehende Auswertung überhaupt erst möglich zu machen.

Die Art der auftretenden Abweichungen ist vielgestaltig. Zum Beispiel denke man an die Abbildungsfehler eines vorgeschalteten Linsensystems, durch die Körnung der Filmbeschichtung bedingtes Rauschen, bewegungsbedingte

Verzerrungen oder, im Falle von Elektronenmikroskopen, Störungen durch Ablenkung. Ebenso zahlreich und vielfältig wie die bei der Bildgewinnung auftretenden Störungen sind daher die Methoden, sie im nachhinein wieder zu beseitigen [Bat86, RK82, Jai81, Literaturhinweise darin].

Dieser Abschnitt soll andeuten, daß Markoff-Felder eine elegante Methode darstellen, eine Vielzahl von Störungen zu modellieren und abzuschwächen. In der Hauptsache wird der vielzitierte Ansatz von Geman & Geman dargestellt [GG84], der sich bahnbrechend für die Verwendung von Markoff-Feldern in der Bildverarbeitung insgesamt auswirkte. Doch nicht nur wegen seiner immensen Bedeutung für die mittlerweile breite Akzeptanz des Markoff-Modells, sondern auch wegen zweier besonderer Techniken im Umgang mit den Markoff-Feldern ist ihm ein Abschnitt gewidmet. Zum einen nämlich wird die Verwendung zusätzlicher Zufallsvariablen zur Beschreibung von abstrakten Bildinhalten wie etwa Kanten vorgeführt, und zum anderen läßt sich gut beobachten, wie die Berechnung einer veränderten Energiefunktion zur Lösung des Problems beitragen kann.

Das Restaurationsproblem besteht darin, ein gegebenes gestörtes Bild in seinen ursprünglichen Zustand vor der Störung zurückzuversetzen. Das hier vorgestellte Verfahren modelliert dazu die möglichen Originalbilder durch ein Gibbs-Zufallsfeld. Zusammen mit der Kenntnis über die vorliegende Störung läßt sich aus dessen Energiefunktion die Wahrscheinlichkeitsverteilung der möglichen Originalbilder bedingt unter dem gegebenen (gestörten) Bild ermitteln. Die Bestimmung eines Bildes, das diese Verteilung maximiert, liefert das – bedingt unter dem aufgezeichneten – wahrscheinlichste ungestörte Bild.

3.2.1 Modifikation des Bildmodells

Wegen der in den meisten natürlichen Szenen vorhandenen scharfen Objektgrenzen erweitern Donald und Stuart Geman das dem Zufallsfeld unterliegende Gitter L_F um ein zweites Gitter L_E, dessen Punkten nicht jetzt nicht mehr Grauwert- sondern Kantenzufallsvariablen zugeordnet werden. Die Kantenelemente liegen horizontal und vertikal zwischen den Pixeln und können zwei Zustände annehmen: Kante vorhanden oder nicht vorhanden. In Abbildung 3.9 ist zu sehen, wie diese beiden Gitter kombiniert werden. Die Elemente des Teilgitters L_F für die Grauwertzufallsvariablen sind durch Quadrate angedeutet, die des Teilgitters L_E für die Kantenzufallsvariablen durch Balken.

Das auf L_F gewählte Nachbarschaftssystem ist das *nearest neighbor* Modell η^1. Auf dem erweiterten Gitter L werden in die Nachbarschaft eines Pixels zusätzlich die zwischen diesem und den Nachbarpixeln liegenden Elemente von L_E aufgenommen. Die Nachbarschaften der Elemente aus L_E bestehen

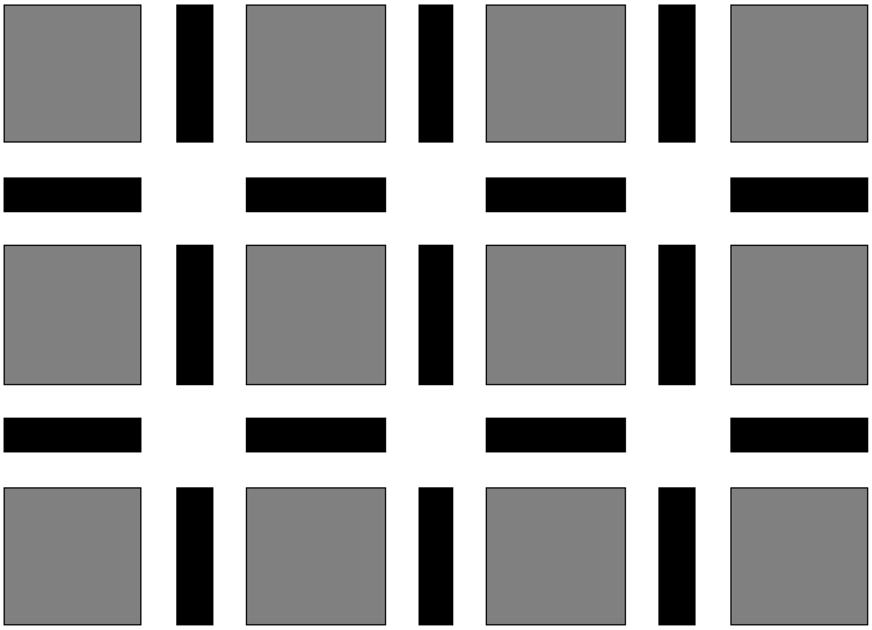


Abbildung 3.9: Ein um Kantenelemente erweitertes Gitter

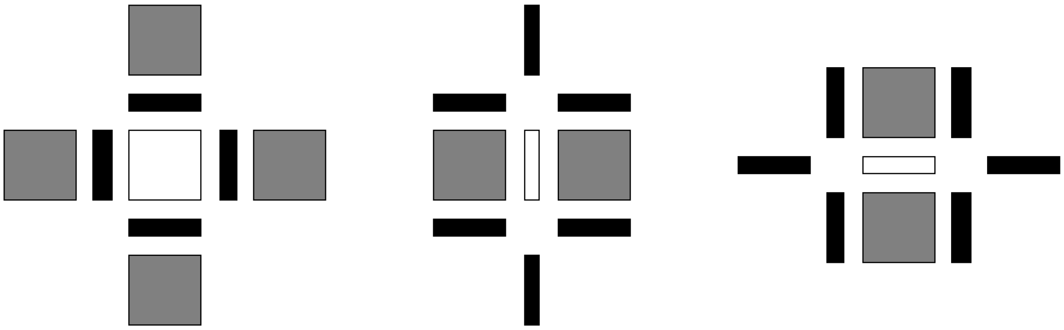


Abbildung 3.10: Nachbarschaften auf $L = L_F \cup L_E$

jeweils aus den beiden nächsten Pixeln und den sechs nächsten Kantenelementen. Sie sind in Abbildung 3.10 dargestellt, Abbildung 3.11 zeigt die daraus enstehenden Cliquentypen.

Das modellierte Bild $x = (f, e) \in \Gamma^{L_F} \times \{0,1\}^{L_E}$ besteht nun aus der Realisierung f des Zufallsfeldes F für das Grauwertbild (*formation*) und der Realisierung e des Zufallsfeldes E (*edges*) für das Kantenbild. In den Komponenten von e stehen Einsen für vorhandene und Nullen für fehlende Kanten.

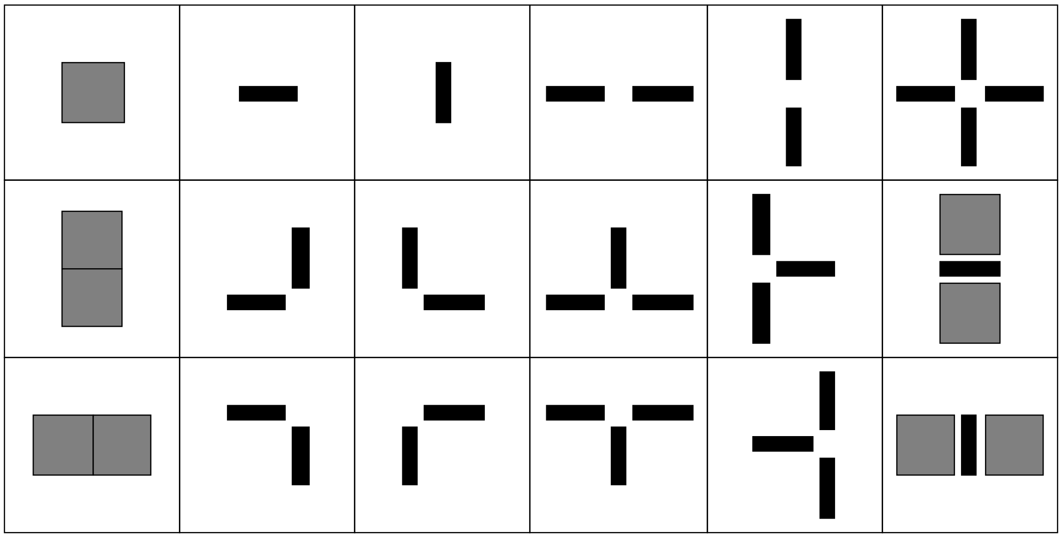


Abbildung 3.11: Cliquentypen auf $L = L_F \cup L_E$

3.2.2 Fehlermodell

Vor der Behebung eines Fehlers muß geklärt werden, worin er besteht. Hier wird der Fall betrachtet, daß das Bild durch ein photochemisches oder photoelektronisches Verfahren entstanden ist. Es ist daher realistisch anzunehmen, daß das Originalbild verschmiert wurde, eine (möglicherweise) nichtlineare Transformation der Grauwerte stattgefunden hat und additives oder multiplikatives Rauschen aufgetreten ist. Die dazugehörigen Rechtfertigungen werden mit den mathematischen Formulierungen der Effekte gegeben.

Verschmierung

Da Lochkameras gewisse praktische Probleme aufwerfen, werden im allgemeinen Aufnahmegeräte mit einem vorgeschalteten Linsensystem verwendet. Dieses bildet ideale Punkte nicht auf ebensolche ab, sondern ist in seinen Auswirkungen durch seine *Point Spread Function* (PSF) charakterisiert. Sie gibt die Antwort an, die das System auf eine Punkteingabe liefert. Geht man davon aus, daß die PSF translationsinvariant ist, so kann die Verzerrung durch Faltung mit einer "Verschmierungsmatrix" $H = (h_{ij})_{\substack{-N+1\leq i\leq N-1\\ -M+1\leq j\leq M-1}}$ ausgedrückt werden:

$$H * F =: H(F)$$

Man beachte, daß H im allgemeinen keine Zufallsvariable ist, sondern als bekannt vorausgesetzt wird. Da die Verschmierungen vorwiegend lokal erfolgen, wird H üblicherweise nur wenige Einträge ungleich Null haben. Als Beispiel möge

$$h_{k,l} = \begin{cases} \frac{1}{2} & k = l = 0 \\ \frac{1}{16} & |k|, |l| \leq 1,\ (k,l) \neq (0,0) \end{cases} \tag{3.3}$$

dienen. Hier wird der Grauwert an einer Stelle (i, j) mit dem arithmetischen Mittel der Grauwerte seiner acht nächsten Nachbarn gemittelt.

Grauwerttransformation

Bedingt durch das technische Verfahren, mit dem das eintreffende Signal abgebildet wird, unterliegen die Intensitäten einer Transformation τ, die in den meisten Fällen nichtlinear ist. Beispielsweise ist die Gradationskurve photochemischer Verfahren weitgehend logarithmisch [Sol71].

Im Grunde ist die Transformation τ hier beliebig wählbar, was gegenüber deterministischen Verfahren, die durch Nichtlinearitäten häufig vor erhebliche Probleme gestellt werden [Hun77, Einleitung], ein wesentlicher Vorteil ist.

Rauschen

Durch thermische Einflüsse in photoelektronischen oder durch molekulare in photochemischen Verfahren [Sol71] der Bildaufzeichnung kommt es fast unvermeidlich zu einer Überlagerung von Rauschen, das additiv oder multiplikativ sein kann. Wie in Beispiel 2.3 wird es durch ein zusätzliches Zufallsfeld $\tilde{N} = (\tilde{N}_{1,1}, \ldots, \tilde{N}_{N,M})$ beschrieben, dessen Komponenten hier als stochastisch unabhängig und identisch verteilt angenommen werden. Konkret sollen sie mit Erwartungswert μ und Varianz σ^2 normalverteilt sein. Außerdem wird vorausgesetzt, daß $\tilde{N}$ stochastisch unabhängig von $X = (F, E)$ ist.

Insgesamt ergibt sich das in Abbildung 3.12 gezeigte Modell für die auftretenden Störungen. Die Zufallsvariable des aufgezeichneten Grauwertbildes $\tilde{F}$ genügt damit der Gleichung

$$\tilde{F} = \tau(H(F)) \oplus \tilde{N} \tag{3.4}$$

oder auf Pixelebene

$$\tilde{F}_{ij} = \tau\left(\sum_{(k,l)\in L} h_{k-i,l-j}\cdot F_{kl}\right)\oplus\tilde{N}.$$

Da es nur auf die Invertierbarkeit der Verknüpfung des verzerrten Bildes mit dem Rauschen ankommt, steht $\oplus$ stellvertretend für Addition oder Multiplikation (oder eine sonstige invertierbare Verknüpfung).

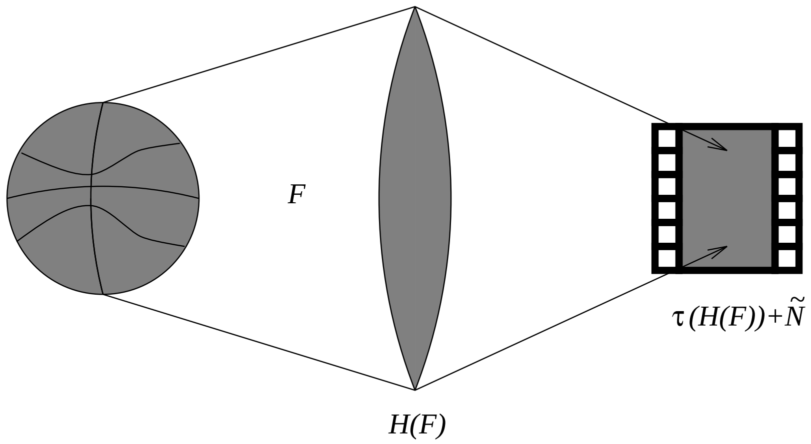


Abbildung 3.12: Skizze zur Bildgewinnung

Die letzte Annahme besteht darin, daß das Kantenbild E nicht verändert wird, das aufgezeichnete, gestörte Bild also beschrieben ist durch $\tilde{X} = (\tilde{F}, E)$.

3.2.3 Maximum a-posteriori Schätzung

Behandelt man Gleichung (3.4) als deterministisches Problem, so ist ein System nichtlinearer Gleichungen zu lösen. Als stochastisches Problem erhält man die Schätzung eines Bildes f, gegeben die Beobachtung $(\tilde{f}, e)$. Mit der Bayes-Formel ergibt sich

$$\begin{aligned} &P(X = (f,e)|\tilde{X} = (\tilde{f},e)) \\ &\quad = \frac{P(\tilde{X} = (\tilde{f},e)|X = (f,e))\cdot P(X = (f,e))}{P(\tilde{X} = (\tilde{f},e))}. \end{aligned} \tag{3.5}$$

Die unter der Beobachtung bedingte Verteilung der möglichen Originalbilder wird *a-posteriori Verteilung* genannt und ist ein vielbenutztes Hilfsmittel der Stochastik [Tan93]. Hat man als Fehlerfunktion die bloße Zählung der Fehleinschätzungen, so führt die Methode der kleinsten Fehlerquadrate (*minimum mean-square error*, MMSE) auf die Bestimmung des Erwartungswerts der *a-posteriori* Verteilung. Die bereits in Abschnitt 3.1.3 aufgetretene Maximum Likelihood Schätzung entspricht der Berechnung eines Maximums von (3.5) unter der Gleichverteilungsannahme für die *a-priori Verteilung* $P(X = (f, e))$.

Die Gleichverteilungsannahme für die *a-priori* Verteilung scheint jedoch nicht angezeigt, da kaum anzunehmen ist, daß jedes beliebige Bild als Ausgangspunkt für die Beobachtung gleich wahrscheinlich ist. Stattdessen wird ein modellbasierter Ansatz verfolgt, d.h. Vorwissen über das Originalbild in Gestalt der *a-priori* Verteilung eingebracht. Anschließende Maximierung von (3.5) bezüglich f wird als *Maximum a-posteriori Schätzung* (MAP) bezeichnet. Da $(\tilde{f}, e)$ das beobachtete Bild ist, kann der Nenner der rechten Seite als konstant angesehen und bei der Maximierung vernachlässigt werden. Das bedeutet insbesondere, daß die meist komplizierte Form der Verteilung $P^{\tilde{X}}$ der gestörten Bilder nicht benötigt wird.

Die MAP Schätzung ist das Kriterium der (optimalen) Bayesschen Entscheidungsregel, und ihre Verwendung ist weit verbreitet [Hun77, Literaturhinweise darin][Tan93]. Geman & Geman [GG84] zeigten nun, daß unter obiger Fehlerannahme mit der *a-priori* auch die *a-posteriori* Verteilung eine Gibbs-Verteilung ist, so daß wegen ihrer speziellen Form (2.9) die Bestimmung ihres Maximums mit der Minimierung der zugehörigen Energiefunktion $\tilde{U}$ zusammenfällt.

3.2.4 A-priori und a-posteriori Verteilung

In diesem Unterabschnitt werden die Annahmen über das Originalbild in der *a-priori* Verteilung codiert und daraus dann die *a-posteriori* Verteilung bestimmt, deren Maximum die Lösung des Schätz- und damit des Restaurationsproblems ist.

Das Originalbild wird als Gibbs-Zufallsfeld über L, bestehend aus L_F und L_E, modelliert. Das unterliegende Nachbarschaftssystem ist η, definiert wie in Abschnitt 3.2.1, und die zugehörige Energiefunktion sei U mit Potential V. Die Wahl der Cliquenpotentiale V_C bestimmt, in welcher Form Vorwissen einfließt (Abschnitt 3.1.2). Sollten keine solchen Annahmen existieren, müssen die Cliquenpotentiale zuvor geschätzt werden (Abschnitt 3.1.3). Die *a-priori* Verteilung hat dann die übliche Form

$$P(F = f, E = e) = \frac{1}{Z(T)}\, e^{-U(f,e)/T}$$

$$Z(T) = \sum_{(f,e)\in\Gamma^{L_F}\cup\{0,1\}^{L_E}} e^{-U(f,e)/T}.$$

Um übersichtlicher notieren zu können, sei o.E. im folgenden $T = 1$. Des weiteren sei η_{ij}^H die Menge der Gitterpunkte aus L_F, die einen Einfluß auf die Verschmierung $H(F)$ im Punkt $(i,j) \in L_F$ haben, und $\eta_{ij}^{H^2}$ sei die in zweiter Ordnung verschmierte Nachbarschaft:

$$\eta_{ij}^{H^2} = \bigcup_{(k,l)\in\eta_{ij}^H} \eta_{kl}^H$$

Sie ist notwendig, um den durch die Verschmierung H neu hinzukommenden Abhängigkeiten von Grauwerten Rechnung zu tragen (siehe auch Beispiel 4.9). Mit diesen Vereinbarungen läßt sich ein Nachbarschaftssystem für die *a-posteriori* Verteilung definieren durch

$$\tilde{\eta}_{ij} = \begin{cases} \eta_{ij} & (i,j) \in L_E \\ \eta_{ij} \cup \eta_{ij}^{H^2} \setminus \{(i,j)\} & (i,j) \in L_F. \end{cases}$$

Man prüft leicht nach, daß es sich bei $\tilde{\eta} = \{\tilde{\eta}_{ij} | (i,j) \in L\}$ tatsächlich um ein Nachbarschaftssystem handelt.

Ist $\ominus$ die zu $\oplus$ inverse Operation, so läßt sich zeigen, daß unter der Fehlerannahme (3.4) die *a-posteriori* Verteilung $P\left(X = (f,e) | \tilde{X} = (\tilde{f}, e)\right)$ für jedes feste $\tilde{f}$ eine Gibbs-Verteilung bezüglich $\tilde{\eta}$ mit Energiefunktion

$$\tilde{U}(f,e) = U(f,e) + \frac{\left\| \overline{\mu} - \left(\tau(H(f)) \ominus \tilde{f}\right)\right\|^2}{2\sigma^2} \tag{3.6}$$

ist [GG84]. Dabei sind $\overline{\mu} = (\mu, \ldots, \mu) \in I\!R^{N \cdot M}$ der Erwartungswertvektor und σ die Varianz des Rauschens $\tilde{N}$.

Durch Minimierung von (3.6) wäre es nun theoretisch möglich, den optimalen Bayes-Schätzer für das Originalbild (f, e) zu ermitteln, was wegen des viel zu großen Suchraums Γ^{L_F} jedoch nicht durchführbar ist. Schon bei kleinen Binärbilder von 64x64 Pixeln hat der Suchraum 2^{4096} Elemente.

Um ein $f \in \Gamma^{L_F}$ zu finden, für das (3.6) möglichst klein wird, führen Geman & Geman eine stochastische Relaxation mit Annealing durch [GG84], die eine Form des Simulated Annealing ist. An dieser Stelle wird darauf allerdings nicht weiter eingegangen, weil in Kapitel 4 ein allgemeines Maschinenmodell entwickelt wird, das Berechnungen der gewünschten Art leistet.

3.2.5 Beispiele

Die Qualität der erhaltenen Schätzungen demonstrieren zwei Beispiele aus [GG84], an denen auch deutlich wird, daß die Erweiterung des Bildmodells um Kantenelemente hilfreich sein kann.

Abbildung 3.13 zeigt ein 64x64 Pixel großes Bild mit 3 Graustufen, d.h. $N = M = 64$ und $G = 3$, und das gleiche Bild nach einer Störung mit additivem Gaußschen Rauschen mit $\mu = 0$ und $\sigma = 0.7$. In der Modellierung der *a-priori* Verteilung wurde als Nachbarschaftssystem zunächst η^2 (ohne zusätzliche Kantenelemente) unterstellt. Die Cliquenpotentiale waren für zweielementige Cliquen $C = \{(i,j),(k,l)\}$ definiert durch

$$V_C(f) = \begin{cases} \frac{1}{3} & f_{ij} = f_{kl} \\ -\frac{1}{3} & f_{ij} \neq f_{kl} \end{cases}$$

und ansonsten identisch Null. Das Ergebnis nach 1000 Iterationen des Optimierungsverfahrens (vgl. Abschnitt 4.2.2) war eher enttäuschend (Abbildung 3.13 c)). Daher wurde das Gittermodell der *a-priori* Verteilung auf die in Unterabschnitt 3.2.1 beschriebene Form gebracht, und die Cliquenpotentiale wurden wie folgt angepaßt: Liegt zwischen zwei Grauwertelementen ein Kantenelement mit Wert Null (keine Kante), so ist das Cliquenpotential wie oben, jedoch mit Wert ± 1 anstelle von $\pm\frac{1}{3}$. Befindet sich zwischen den beiden Pixeln eine Kante, so wird die Bindung aufgehoben und das Cliquenpotential auf Null gesetzt. Von den reinen Kantencliquen erhalten nur vierelementige ein Potential ungleich Null. Die einzelnen Werte der sechs möglichen Typen sind in Abbildung 3.14 aufgeführt. Abbildung 3.13 d) zeigt das deutlich bessere Ergebnis nach ebenfalls 1000 Iterationen.

Man erkennt leicht, daß die Wahl der *a-priori* Verteilung wesentlichen Einfluß auf die Qualität der Restauration hat. Einerseits ist damit eine gute Möglichkeit gegeben, Vorwissen oder Annahmen in das Verfahren zu integrieren. Andererseits birgt dies natürlich auch eine Quelle für unnötige Fehler. Wünschenswert wäre es daher, die *a-priori* Verteilung möglichst gut aus den vorhandenen Bildinformationen zu schätzen.

Im zweiten Beispiel wurde dasselbe Originalbild mit der Matrix (3.3) verschmiert und mit $\tau(g) = \sqrt{g}$ transformiert. Anschließend wurde multiplikatives Rauschen mit $\mu = 1$ und $\sigma = 0.1$ hinzugefügt. Die Restauration erfolgte mit der gleichen *a-priori* Verteilungsannahme wie im zweiten Fall des ersten Beispiels mit wiederum 1000 Iterationen (Abb. 3.15).

3.2.6 Einige weitere Ansätze

Der zitierte Artikel der Geman-Brüder war nicht der erste Ansatz zur Bildrestauration mit Hilfe von Markoff-Feldern. Beispielsweise wurden sie bereits ein Jahrzehnt zuvor von Jain und Angel [JA74] benutzt, um rekursive Filter für die Restauration verrauschter Bilder zu konstruieren.

Dennoch wirkte die zitierte Arbeit von Geman & Geman bahnbrechend für den Einsatz räumlicher Wechselwirkungsmodelle, und das nicht nur bei der

a) **b)**

c) **d)**

Abbildung 3.13: Auswirkungen der Kantenelemente
a) Originalbild, b) gestörtes Bild: additives Rauschen, c) Restauration ohne Kantenelemente, d) Restauration mit Kantenelementen. Aus [GG84].

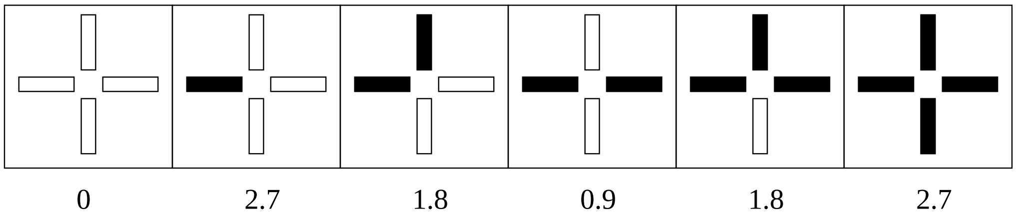


Abbildung 3.14: Cliquenpotentiale der vierelementigen Cliquen des Kantengitters (um 90^o, 180^o bzw. 270^o gedrehte Konfigurationen werden gleich bewertet).

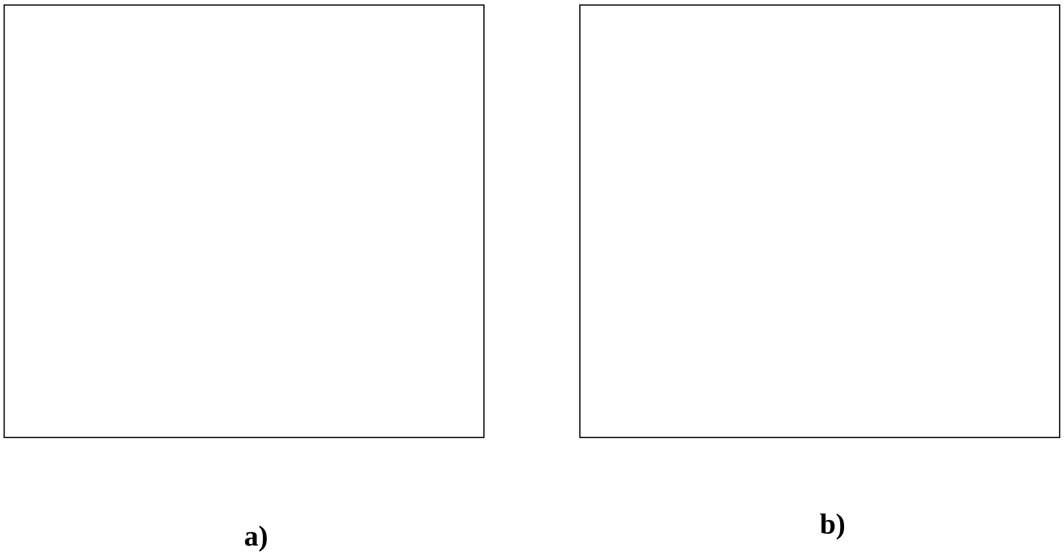


Abbildung 3.15: Bildrestauration

a) Gestörtes Bild: Verschmierung, Grauwerttransformation, multiplikatives Rauschen, b) Restauration (mit Kantenelementen). Aus [GG84].

Restauration, sondern auch in vielen anderen Gebieten der Bildverarbeitung. Naturgemäß haben dabei zahlreiche Autoren das oben dargestellte Konzept aufgegriffen und abgewandelt oder erweitert.

So verwendet Chalmond [Cha88] eine *a-priori* Verteilung mit parametrisierter Energie und schätzt die Parameter aus dem beobachteten Bild, bevor die eigentliche Restauration in analoger Weise ansetzt. Lediglich das Abkühlschema für die Relaxation wird gegenüber dem Algorithmus in [GG84] modifiziert.

Wenig später stellte Chalmond [Cha89] einen Ansatz vor, der die Parameterschätzung für die Energie der *a-priori* Verteilung in einem iterativen Verfahren mit der Restauration abwechselnd leistet und auf dem in Abschnitt 3.4 näher erläuterten EM-Algorithmus basiert. Anstelle eines festen (gewählten oder geschätzten) Modells der *a-priori* Verteilung wird also eine Vermutung über ihr Aussehen immer wieder an das sich langsam verbessernde Bild angepaßt.

Im Fall von Mehrpegelbildern, die nur wenige, weit über den Grauwertbereich Γ verteilte Grauwerte (Pegel) enthalten, fallen Restauration und Segmentierung von lediglich verrauschten Bildern zusammen, denn die Regionen gleichen Grauwertes bilden die Segmente. Shridar, Ahmadi und El-Gabali [SAG89] geben drei Verfahren zur Bestimmung einer MAP-Schätzung an, die jedoch nicht ohne die in mancher Hinsicht aufschlußreichen Kommentare von Derin [Der91] hingenommen werden sollten. Ein von Lakshmanan und Derin [LD89] veröffentlichtes Verfahren segmentiert verrauschte Bilder und schätzt, ähnlich wie bei Chalmond [Cha89], die Parameter simultan. Allerdings basiert ihr Algorithmus wieder auf dem bereits angesprochenen und in Kapitel 4 behandelten Simulated Annealing für Zufallsfelder.

3.3 Textursegmentierung

Es wurde bereits festgestellt, daß das Segmentierungsproblem in wichtigen Spezialfällen mit dem Restaurationsproblem zusammenfällt. In diesem Abschnitt geht es daher um eine andere Art von Segmentierung, bei der das Kriterium für den Zusammenhang einer Bildregion nicht der vorherrschende Grauwert (Pegel), sondern eine Textur ist. Das Segmentierungsproblem wird dadurch ungleich schwieriger und bietet eine anschauliche Möglichkeit, eine besonders interessante Variante des stochastischen Bildmodells vorzuführen.

Die Darstellung folgt dem von Derin und Elliott [DE87] vorgeschlagenen Verfahren. Obwohl daraus wiederum eine Lösung durch Bestimmung eines MAP-Schätzers resultiert, werden zwei ganz anders geartete Wege zur Modellierung und Optimierung beschritten, als es in Abschnitt 3.2 im Zusammenhang mit der Bildrestauration der Fall war.

Konkret wird anstelle der Kombination mehrerer Felder für verschiedene Merkmale (wie eben Grauwerte und Kanten) eine Hierarchie von Feldern als Mittel zur Beschreibung unterschiedlicher Abstraktionsgrade eingeführt. Des weiteren wird anstelle der stochastischen Approximation ein exaktes deterministisches Verfahren, die Dynamische Programmierung[2], zur Schätzung eingesetzt. Aufgrund des hohen Speicherbedarfs muß es allerdings modifiziert werden, weswegen das Ergebnis trotzdem nur eine Näherungslösung verkörpert.

3.3.1 Problemstellung

Viele Anwendungen der Bildverarbeitung machen es notwendig, das erhaltene Bild in "homogene" Regionen zu zerlegen – es zu segmentieren. Diese Regionen können einzelne Objekte oder sonstige, durch ein einheitliches Erscheinungsbild ausgezeichnete Bildteile sein. Im einfachsten Fall besteht jede Region aus einer Menge von Pixeln des gleichen Grauwerts. Ein solches Bild soll als bereits segmentiert angesehen werden, weil aufgrund der Grauwerte sofort über die Zugehörigkeit zu einer Region entschieden werden kann.

Tatsächlich wird man es aber mit Bildern wie dem in Abbildung 3.16 gezeigten zu tun haben. Obwohl die Regionen gleichen Grauwerts andere sind, wird es dem menschlichen Betrachter sicherlich nicht allzu schwerfallen, beispielsweise die Knochenpartien vom Hintergrund zu unterscheiden. Der Grund dafür liegt in der weitgehend einheitlichen Erscheinung der jeweiligen Regionen. Dieses "einheitliche Erscheinungsbild" läßt sich als Textur auffassen.

Das Textursegmentierungsproblem besteht demnach darin, in einem gegebenen Bild die Regionen gleicher Textur zu ermitteln. Aus dem hier zur Lösung verwendeten Bildmodell wird sich eine elegante Formalisierung dieser Problemstellung ergeben.

3.3.2 Ein hierarchisches Bildmodell

Um die Zusammensetzung eines Bildes aus Regionen gleicher Textur zu modellieren, wird ein hierarchisches Modell benutzt. Die im Bild auftretenden Texturen werden durch Markoff-Felder T^t, $t \in \{1, \ldots, \Theta\}$ beschrieben und bilden die untere Modellebene. Ein weiteres, übergeordnetes Markoff-Feld X

[2] Auf eine Erläuterung des allgemeinen Ansatzes wird an dieser Stelle verzichtet. Sie kann z.B. in [Aig93] nachgelesen werden.

Abbildung 3.16: Röntgenbild
Röntgenaufnahme eines menschlichen Knies mit Knochentumor (aus [Wei93]).

bestimmt, welche Textur an einem gegebenen Bildpunkt vorliegt. Es wird die *Szene* genannt, weil es von den eigentlichen Grauwerten abstrahiert und zur gleichen Textur gehörige Pixel nicht unterscheidet. Zusammengenommen ergibt sich für die Komponenten Y_{ij} des Zufallsfeldes Y für das texturierte Bild

$$Y_{ij} = T_{ij}^{X_{ij}}$$

oder, anders ausgedrückt,

$$Y_{ij}(\omega) = T_{ij}^{t}(\omega), \text{ falls } X_{ij}(\omega) = t$$

für alle $\omega \in \Omega$, dem Ergebnisraum des zugrundeliegenden Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, P)$.

Beiden Ebenen des Modells unterliegt das gleiche Gitter L mit Nachbarschaftssystem η, und es wird $\eta = \eta^2$ angenommen. Die Wahl von η^2 ist willkürlich und wurde lediglich aus praktischen Gesichtspunkten heraus getroffen, um den Rechenaufwand in Grenzen zu halten. Theoretisch ließe sich der Algorithmus jedoch leicht an andere Nachbarschaftssysteme anpassen. Die Texturen realisieren sämtlich in der Grauwertmenge Γ, das übergeordnete Feld X dagegen realisiert in $\{1, \ldots, \Theta\}$, der Menge der Texturindizes.

Man kann sich Y als zweistufiges Verfahren vorstellen: Das Markoff-Feld X bestimmt, welcher Texturtyp t an einem Punkt vorliegt, und das Feld T^t entscheidet dann jeweils, mit welchem Grauwert dieses Pixel eingefärbt wird. Abbildung 3.17 veranschaulicht diese beiden Stufen.

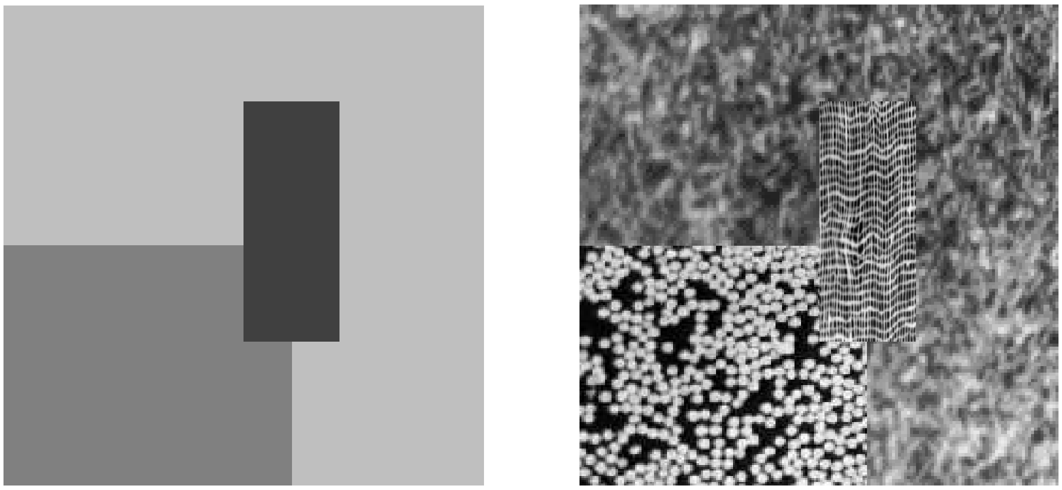


Abbildung 3.17: Hierarchisches Markoff-Feld

Links die *Szene* X (Texturindizes $t \in \{1, 2, 3\}$ durch Grauwerte dargestellt), rechts die gleiche Szene mit texturierten Regionen. Texturen aus [Bro66].

In Abbildung 3.17 erkennt man aber auch sofort, daß es sich bei der Szene X um die Lösung des Segmentierungsproblems $Y = T^X$ handelt.

Um aus einem texturierten Bild y die vermutliche Szene x zu bestimmen, aus der y durch Texturierung enstanden ist, wird wiederum der Maximum *a-posteriori* Schätzer (MAP, vgl. Abschnitt 3.2.3) bemüht, also ein x gesucht, das für festes y

$$P\left(X = x | Y = y\right) = \frac{P\left(Y = y | X = x\right) \cdot P\left(X = x\right)}{P\left(Y = y\right)} \tag{3.7}$$

bzw., äquivalent dazu,

$$\log P(X = x, Y = y) = \log P(X = x) + \log P(Y = y | X = x) \tag{3.8}$$

maximiert.

3.3.3 MAP-Schätzung durch Dynamische Programmierung

Zunächst soll gezeigt werden, daß sich das durch (3.8) gegebene Maximierungsproblem unter einigen zusätzlichen Annahmen exakt lösen läßt. Weil dazu jedoch ein untragbarer Aufwand erforderlich wäre, wird anschließend eine vertretbare Einschränkung vorgenommen, durch die dann natürlich keine Garantie mehr für die Optimalität der Lösung besteht.

Vorbereitung

Um das Prinzip der Dynamischen Programmierung anwenden zu können, wird die rechte Seite von (3.8) in eine Summe entwickelt. Für den ersten Summanden, die logarithmierte Verteilung der Szene, gilt

$$\begin{aligned}\log P(X=x) &= -\log Z(T) - \sum_{C\in\mathcal{C}} V_C(x)/T \\ &= -\log Z(T) - \sum_{i=1}^{N} \sum_{C\in\mathcal{C}^{i-1,i}} V_C(x)/T.\end{aligned}$$

Dabei ist $\mathcal{C}^{i-1,i} \subseteq \mathcal{C}$ die Menge der Cliquen, die nur Punkte aus der i-ten Spalte oder Punkte aus den Spalten $i-1$ und i enthalten, also

$$C \in \mathcal{C}^{i-1,i} \iff \begin{cases} & C \in \mathcal{C} \\ \wedge & C \subseteq \{i-1, i\} \times M \\ \wedge & C \cap \{i\} \times M \neq \emptyset. \end{cases}$$

Die zweite Gleichheit der Summenentwicklung gilt dann, weil wegen $\eta = \eta^2$ keine drei- oder mehrspaltigen Cliquen vorkommen.

Der letzte Ausdruck für $\log P(X = x)$ läßt sich in rekursiver Form notieren:

$$\begin{aligned} l'_0 &= -\log Z(T) \\ l'_i &= l'_{i-1} - \sum_{C\in\mathcal{C}^{i-1,i}} V_C(x)/T \qquad i = 1,\ldots,N \end{aligned} \tag{3.9}$$

Es gilt dann $l'_N = \log P(X = x)$, womit allerdings erst ein Summand für die Dynamische Programmierung vorbereitet wäre. Um auch den zweiten in eine Summe zu entwickeln, werden drei zusätzliche Annahmen benötigt:

$$\begin{aligned} (i) \qquad P(X=x) &= \prod_{(i,j)\in L} P(X_{ij} = x_{ij}) \\ (ii) \qquad P(X=x|Y=y) &= \prod_{(i,j)\in L} P(X_{ij} = x_{ij}|Y=y) \\ (iii) \quad P(X_{ij}=x_{ij}|Y=y) &= P(X_{ij} = x_{ij}|Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}}) \end{aligned} \tag{3.10}$$

Obwohl diese Annahmen im allgemeinen nicht gelten, lassen sie sich wie folgt rechtfertigen [DE87]: Da die Wechselwirkungen des Szene-Modells X bereits im ersten Summanden von (3.8) berücksichtigt sind, ist es vertretbar zu unterstellen, die X_{ij} seien stochastisch unabhängig und bedingt stochastisch unabhängig gegeben $Y = y$. Die dritte Annahme dagegen erscheint ohnehin vernünftig, denn sie besagt lediglich, daß die Region (Textur), zu der ein Punkt gehört, schon durch seinen und die Grauwerte der Nachbarpunkte festgelegt ist.

Setzt man nun diese Annahmen in (3.7) ein, so ergibt sich

$$\begin{aligned} P(Y = y|X = x) &= \frac{\prod_{(i,j)\in L} P(X_{ij} = x_{ij}|Y = y)P(Y = y)}{P(X = x)} \\ &= P(Y = y) \prod_{(i,j)\in L} \frac{P(X_{ij} = x_{ij}|Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}})}{P(X_{ij} = x_{ij})} \\ &= \frac{P(Y = y)}{\prod_{(i,j)\in L} P(Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}})} \prod_{(i,j)\in L} P(Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}}|X_{ij} = x_{ij}). \end{aligned}$$

Der erste Faktor des letzten Ausdrucks ist von x unabhängig und kann daher als Konstante c_0 behandelt werden. Mit der Definition $S_t^i := \{i\} \times \{j \in M|X_{ij} = t\}$ kann der zweite Summand aus (3.8) nun in der Form

$$\log P(Y = y|X = x) = \log c_0 + \sum_{t=1}^{\Theta} \sum_{i=1}^{N} \sum_{(i,j)\in S_t^i} \log P(Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}}|X_{ij} = t)$$

notiert werden. Zusammen mit (3.9) erhält man für die logarithmierte gemeinsame Verteilung von Szene und texturiertem Bild die Rekursion

$$\begin{aligned} l_0 &= l_0' + \log c_0 = -\log Z(T) + \log c_0 \\ l_i &= l_{i-1} - \sum_{C\in\mathcal{C}^{i-1,i}} V_C(x)/T \\ &\quad + \sum_{t=1}^{\Theta} \sum_{(i,j)\in S_t^i} \log P(Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}}|X_{ij} = t) \qquad i = 1, \ldots, N. \end{aligned} \tag{3.11}$$

Unter den Annahmen (3.10) gilt also insgesamt $l_N = \log P(X = x, Y = y)$.

Modifikation der Dynamischen Programmierung

Die Rekursion (3.11) erfüllt offensichtlich das Optimalitätskriterium von Bellmann, so daß das Prinzip der Dynamischen Programmierung zur Maximierung von l_N anwendbar ist. Da l_0 von keiner Segmentierung x abhängt, kann darüber hinaus $l_0 = 0$ gesetzt werden, was die schwierige Berechnung von $Z(T)$ vermeidet. Des weiteren stellt die explizite Berechnung der Wahrscheinlichkeiten $P(Y_{\overline{\eta_{ij}}} = y_{\overline{\eta_{ij}}}|X_{ij} = t)$ kein allzu großes Problem dar, weil die Nachbarschaften in η^2 relativ klein sind.

Nun gibt es zu jeder der N Spalten eines Bildes aber Θ^M mögliche Segmentierungen, so daß der Algorithmus einen Aufwand von $\mathcal{O}(N \cdot \Theta^{2M})$ Schritten erfordert und damit ab etwa $M > 4$ nicht mehr durchführbar ist. Aus diesem Grunde muß man auf eine optimale Lösung verzichten und sich mit einer Näherung begnügen. Das hier verwendete Prinzip wurde schon vorher zur Segmentierung verrauschter Bilder benutzt [DECG84] und fixiert weiter zurückliegende Entscheidungen der Dynamischen Programmierung.

Die modifizierte Version des Algorithmus' ergibt sich aus der sukzessiven Bearbeitung von relativ schmalen horizontalen Streifen. Die Rekursion wird in mehreren Durchgängen jeweils nur auf $D \leq 4$ Zeilen des Bildes angewandt. Beginnnt man am oberen Bildrand, so wird l_N für das auf die D ersten Zeilen eingeschränkte Bild maximiert und von der erhaltenen Schätzung nur die erste Zeile gespeichert. Anschließend wird die Berechnung in gleicher Weise für die Zeilen 2 bis $D + 1$ wiederholt, wobei die Schätzung (Segmentierung) der ersten Zeile als Randbedingung dient. Vom zweiten Durchlauf wird wieder die oberste Zeile verwahrt und der Vorgang eine Zeile tiefer erneut gestartet. Die Berechnung endet mit dem Durchlauf für die D untersten Zeilen, deren komplette Schätzung behalten werden muß. Fügt man schließlich die geschätzten Zeilen aneinander, erhält man eine Näherung der optimalen Segmentierung.

Die beiden verschachtelten Teile des Algorithmus können als doppelte Optimierung angesehen werden: Eine horizontale, exakte, und eine vertikale, näherungsweise. Wie ein Fenster wird ein D Zeilen breiter Streifen von oben über das Bild geschoben und horizontal die Rekursion (3.11) durchlaufen, wobei jeweils eine Zeile der Segmentierung geschätzt wird.

Die zur Modifikation gewählte Einschränkung wird plausibel [DE87], wenn man berücksichtigt, daß immer nur der Teil der Segmentierung festgehalten wird, der von den noch nicht bearbeiteten Daten (den Zeilen weiter unten) am weitesten entfernt ist. Dahinter steht die Annahme, daß die D Zeilen auseinanderliegenden Zufallsvariablen der Szene vernachlässigbare Abhängigkeiten aufweisen.

3.3.4 Beispiel und Anmerkungen

Abbildung 3.18 zeigt ein Textursegmentierungsproblem, das recht gut gelöst wird. Die drei Regionen der Szene wurden mit Texturen überzogen und anschließend mit dem oben beschriebenen Algorithmus zurückgewonnen. Das Differenzbild zeigt, daß die meisten Fehler erwartungsgemäß am Rand der Regionen aufgetreten sind. Man beachte aber, daß das Modell keine Störungen vorsieht, so daß nur unverrauschte Ausgangsbilder zugelassen sind.

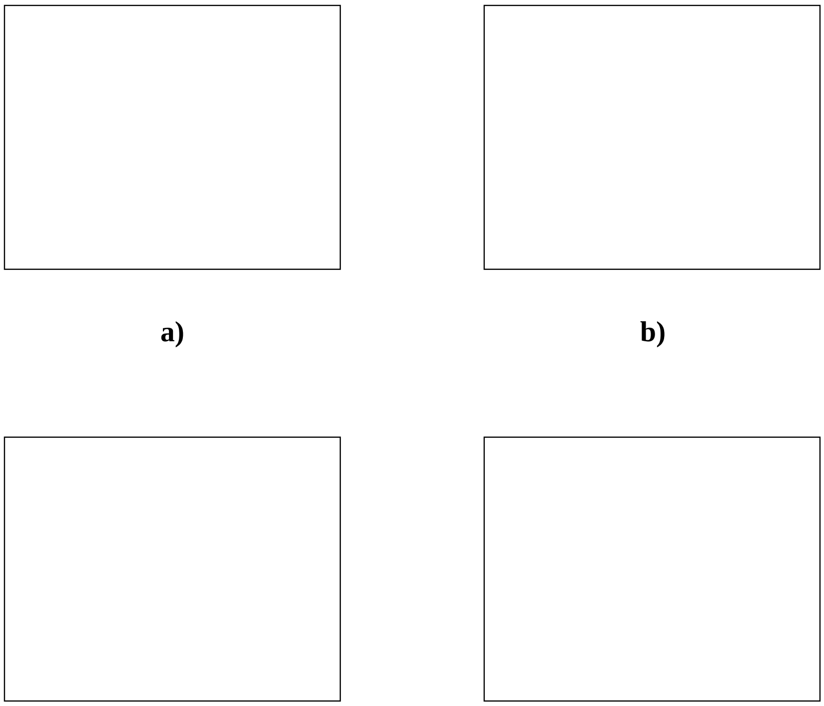


Abbildung 3.18: Segmentierungsbeispiel

Segmentierung eines Bildes mit drei Regionen und vier Grauwerten: a) Szene, b) texturiertes Bild, c) Segmentierung, d) Differenzbild (Fehlerrate 2.3%). Aus [DE87].

Die Autoren berichten, daß die Wahl von $D = 3$ bei allen Tests zufriedenstellende Ergebnisse lieferte, die durch $D = 4$ nicht wesentlich verbessert werden konnten [DE87]. Die Vermutung, das Nachbarschaftssystem η^2 sei der Hauptgrund dafür, liegt nahe. Zahlreiche weitere Beispiele für die Anwendung des behandelten Verfahrens werden in [DC86] gegeben.

Problematisch ist allerdings, daß die von Derin und Cole [DC86] bzw. Derin und Elliott [DE87] angeführten Beispiele mit vorgegebenen Parametern künstlich generiert wurden. In wirklichen Anwendungen wird man nicht davon ausgehen können, das Nachbarschaftssystem und die Cliquenpotentiale von Szene und Texturen zu kennen. Wiederum tritt damit das Problem der Parameterschätzung auf, das ganz wesentlich in die Leistungsfähigkeit und Anwendbarkeit der Algorithmen eingeht (siehe auch Abschnitte 3.1.3 und 4.5).

Es muß ferner darauf hingewiesen werden, daß der Aufwand für einen Durchlauf, also die Optimierung eines Streifens der Breite D, größenordnungsmäßig immerhin noch $\mathcal{O}(N \cdot \Theta^{2D})$ Schritte beträgt. Mit anderen Worten wird auch das eingeschränkte Verfahren für Bilder mit mehr als 4 Regionen zu aufwendig. Derin und Elliott schlagen für $\Theta > 4$ einen hierarchischen Algorithmus aus Zerlegungen in jeweils zwei Regionen vor [DE87]. Ein solcher Binärzerlegungsansatz wird bei Shridar *et al.* [SAG89, Referenzen darin] ausführlicher behandelt.

Für den Algorithmus spricht andererseits, daß er das Bild nur ein einziges Mal abarbeitet. Bei den stochastischen Optimierungsverfahren zur Schätzerbestimmung ist die Laufzeit schwierig vorherzusagen, da die Konvergenzgeschwindigkeit nicht offensichtlich ist. Bei diesem Verfahren aber steht die Laufzeit von vorneherein fest. Und in Anbetracht des Schwierigkeitsgrades der Aufgabenstellung sind die Ergebnisse zufriedenstellend.

3.4 Bildrekonstruktion

Als letzte soll eine Anwendung aus der medizinischen Diagnostik ausführlicher behandelt werden. Bildgebende Verfahren sind für die Medizin ein unverzichtbares Hilfsmittel und Gegenstand intensiver Forschung. Seit der Entdeckung der bis dahin unbekannten elektromagnetischen Strahlung im Wellenlängenbereich unter 10 nm durch Röntgen (1895) wurden die Methoden und Apparate der Radiologie unablässig weiterentwickelt [TB86].

Die durch Hounsfield 1971 eingeführte Computertomographie (CT) ist ohne die Verwendung von geeigneten Algorithmen für die Rekonstruktion der räumlichen Information aus den Projektionen gar nicht erst möglich. Somit ist erstmals ein bildgebendes Verfahren entwickelt worden, das eine Verarbeitung der anfallenden Daten erzwingt. Bei der Emissions-Computertomographie (E-CT) kommen für diese Bearbeitung erschwerend die statistischen Merkmale des radioaktiven Zerfalls hinzu, so daß der Einsatz stochastischer Methoden zumindest naheliegt.

Auch bei dem hier für die Darstellung ausgewählten Ansatz [Gre90] für die Single-Photon-ECT (SPECT) besteht der methodische Beitrag in einem wichtigen Verfahren zur Bestimmung eines MAP-Schätzers (hier für Schnittbilder). Es zeichnet sich durch den Umgang mit "unvollständigen" Daten (den Projektionen) aus und ist in der Stochastik als EM-Algorithmus [DLR77] bekannt.

3.4.1 SPECT

In der Nuklearmedizin werden radioaktiv markierte Stoffe unter anderem zu diagnostischen Zwecken eingesetzt. Radioaktivität wiederum ist die spontane Umwandlung von Atomkernen unter Aussendung von Strahlen, von denen hier besonders Gamma- und Röntgenstrahlen interessieren [Win90]. Instabile Atomkernarten, die von selbst eine Kernumwandlung durchmachen, heißen Radionuklide. Der stattfindende Umwandlungsprozeß ist von außen nicht zu beeinflussen, sondern rein statistischer Natur und für die jeweilige Umwandlung spezifisch [Lis86].

In der Radiologie wird die Gammastrahlung ausgenutzt, indem dem Patienten ein mit Radionukliden angereichertes Pharmazeutikum (sog. Radiopharmazeutikum) verabreicht wird, das sich in dem interessierenden Organ auf eine Weise ansammelt, die Aufschluß über zu untersuchende Stoffwechselvorgänge liefert. Die bei SPECT typischerweise verwendeten Radionuklide sind (je nach untersuchter Körperregion) ^{99m}Tc, ^{123}J, ^{131}J, ^{67}Ga und ^{201}Tl. Beispielsweise kommen bei Untersuchungen des myokardialen Stoffwechsels mit ^{123}J angereicherte Fettsäuren zum Einsatz [Win90].

Mit Hilfe eines Detektorsystems, das um eine längs durch den Körper des Patienten verlaufende Achse rotiert wird, werden die in der jeweiligen Richtung emittierten Photonen gezählt. Die schematische Anordnung ist in Abbildung 3.19 dargestellt, Beispiele für resultierende Projektionen zeigt Abbildung 3.20 a)–c). Aus den Zählungen ist dann die räumliche Konzentrationsverteilung des Pharmazeutikums zu rekonstruieren. Für die in dieser Arbeit verfolgten Zwecke reicht es aus, die Gewinnung eines Schichtbildes zu demonstrieren, weswegen jeweils nur die von einer einzelnen Detektorreihe aufgenommenen Projektionen betrachtet zu werden brauchen.

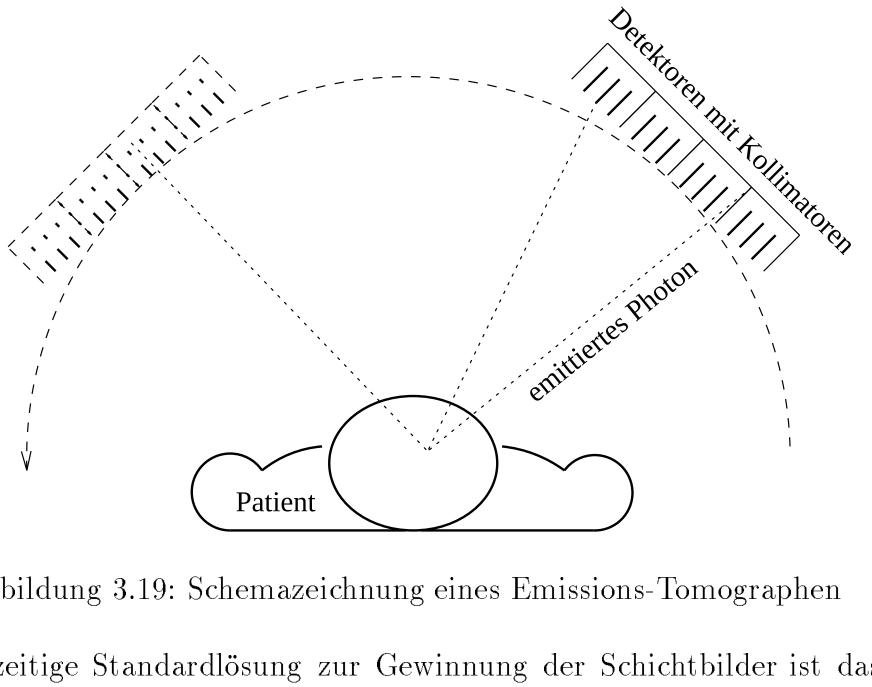


Abbildung 3.19: Schemazeichnung eines Emissions-Tomographen

Die derzeitige Standardlösung zur Gewinnung der Schichtbilder ist das lineare Verfahren der gefilterten Rückprojektion (*filtered back-projection*, z.B.

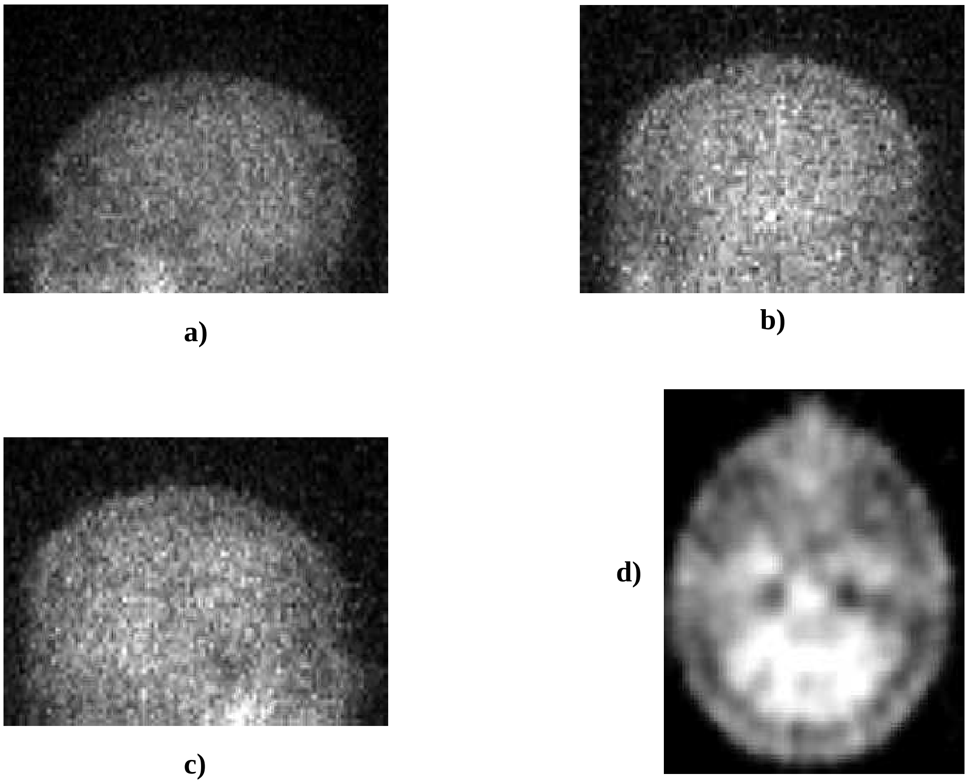


Abbildung 3.20: Beispiel für eine SPECT-Aufnahme
a)–c) Projektionen unterschiedlichen Aufnahmewinkels, d) Schichtbildrekonstruktion mittels gefilterter Rückprojektion. Zur Verfügung gestellt vom Forschungszentrum Jülich.

[RK82, Bat86]), welches jedoch zu wohlbekannten Abbildungsfehlern führt (in erster Linie radiale Artefakte, Abb. 3.20 d)). Green [Gre90] zufolge besteht der zentrale Ansatz zur Verbesserung der Ergebnisse in der genaueren Modellierung des Emissionsvorgangs.

3.4.2 Modellierung des Abbildungsprozesses

Es sind zum einen die tatsächliche Aktivitätsverteilung im Körper des Patienten und zum anderen die gemessenen Photonenemissionen zu modellieren.

Der Patientenkörper wird dazu in kleine Quader eingeteilt, über welche die Zerfallszahlen jeweils summiert werden. Durch die Beschränkung auf einzelne Schichten reicht eine zweidimensionale Indizierung der Quader aus. Die betrachtete Ebene des Körpermodells kann somit als Gitter L beschrieben werden, wobei jeder Gitterpunkt einem Quader im zugehörigen Querschnitt des Patientenkörpers entspricht. Der Einfachheit halber wird wieder Pixel (statt z.B. *Voxel*) gesagt, wenn ein solches Raumelement gemeint ist.

Die (unbekannte) Radionuklidkonzentration in der Querschnittsebene wird durch ein Zufallsfeld X über L modelliert. Die Grauwerte aus Γ sollen unmittelbar mit der Radioaktivität korrespondieren.

Die üblicherweise als Detektor verwandte Gamma-Kamera zählt wegen des davor befindlichen Kollimators nur solche Photonen, deren Pfad nahezu senkrecht zur Aufnahmefläche verläuft. Die Zahl der detektierten Photonen wird für eine Projektion festgehalten und das Detektorarray ein Stück weit um den Körper des Patienten rotiert. Die gemessenen Anzahlen, hier *Zählungen* genannt, sind daher indiziert mit dem Winkel der Projektion und der Position des Detektors im Array. Ohne Einschränkung wird jede Zählung y_φ nur mit einem Winkel φ indiziert und so getan, als gebe es zu jedem dieser Winkel einen fest installierten Detektor. Φ bezeichne die Menge aller Detektoren (Projektionswinkel) und Y sei ein Zufallsvektor mit Komponenten Y_φ zur Beschreibung der Zählungen.

Zusammen sind damit definiert ein "Originalbild" X und eine Beobachtung Y, aus der X zu rekonstruieren ist. Green verwendet hierzu ebenfalls einen MAP-Schätzer, für den aufgrund von (3.7) wie üblich zwei Verteilungen benötigt werden, nämlich zum einen die Likelihood $P^{Y|X}$, die hier das Aufzeichnungsmodell repräsentiert, und zum anderen die *a-priori* Verteilung P^X.

Das Aufzeichnungsmodell

Es wird unterstellt, daß die Photonemissionen eines Pixels (Quaders) vollständig durch eine zeitlich konstante Zerfallsrate beschrieben sind. Die Anzahl x_{ij} der vom Pixel (i,j) emittierten und in Detektor φ ankommenden Photonen bildet dann einen Poisson-Prozeß. Jedes Photon wird höchstens einmal gezählt, und die jeweiligen Poisson-Prozesse sind voneinander unabhängig. Die Gesamtzahl y_φ der in Detektor φ ankommenden Photonen ist damit ebenfalls Poisson-verteilt, und zwar mit Parameter

$$\lambda_\varphi = \sum_{(i,j)\in L} \gamma_{ij,\varphi} x_{ij}.$$

Dabei ist $\gamma_{ij,\varphi}$ ein Maß für die Wahrscheinlichkeit, daß von Pixel (i,j) ausgesandte Photonen in Detektor φ registriert werden. Vor einer näheren Erläuterung kann aber schon festhalten werden, daß sich wegen der Unabhängigkeit der Zählungen in den Detektoren die Likelihood einer Beobachtung y unter der Konzentration x zu

$$P(Y=y|X=x) = \prod_{\varphi\in\Phi} \frac{(\lambda_\varphi)^{y_\varphi}}{y_\varphi!}\, e^{-\lambda_\varphi}$$

berechnet.

Doch zurück zu den Gewichten $\gamma_{ij,\varphi}$. Die Frage, wie viele in Pixel (i,j) ausgesandte Photonen in Detektor φ registriert werden, hängt von zahlreichen Faktoren ab. Da die Halbwertszeit durch multiplikative Anpassung der Gewichte während der Untersuchung berücksichtigt werden kann, geht der neben der Streuung der Photonen wichtigste Einfluß von der Absorption im Körper aus. Die bei SPECT auftretende Strahlung hat je nach verwendetem Radionuklid eine Energie von 70 bis 360 keV, so daß die Absorption auch dünnerer Weichteilschichten nicht zu vernachlässigen ist [Lis86, Seite 8]. Da der Grad der Absorption je nach Medium (Knochen, Gewebe, Luft) stark variert, wäre im Prinzip über die verschiedenen Absorptionsraten entlang des Ausbreitungsweges zu integrieren.

Es ist zwar wichtig, sich klarzumachen, daß hier ein essentielles Problem vorliegt, da es für diese Arbeit aber ohne Belang ist, werden die Schwierigkeiten in der Modellierung der Gewichte übergangen, und es wird auf den zitierten Artikel Greens sowie die darin enthaltenen Anmerkungen verwiesen.

Die a-priori Verteilung

Wie in den zuvor behandelten Anwendungen auch läßt das Problem der *a-priori* Verteilung zwei Sichtweisen zu. Einerseits bietet sich die Möglichkeit, Vorwissen über Form und Größe des zu untersuchenden Organs in die Cliquenpotentiale einer Gibbs-Verteilung einzubringen. Andererseits kann dieser scheinbare Vorteil des Verfahrens auch als Nachteil angesehen werden, weil damit die Güte der Rekonstruktion beträchtlich von nicht immer leicht zu rechtfertigenden Parametern abhängt. Wiederum wäre also eine gute Parameterschätzung wünschenswert. Allerdings kompliziert sie sich im vorliegenden Fall noch erheblich, weil anstelle eines (gestörten) Schichtbildes lediglich Projektionen desselben beobachtet werden.

Green wählt als *a-priori* Verteilung eine Klasse von Gibbs-Verteilungen, die zu Auto-Modellen mit Nachbarschaftssystem η^2 führen, bei denen nur Zweiercliquen ein Potential ungleich Null erhalten. Das Potential der Zweiercliquen ist parametrisiert und hat die Form

$$V_{\{(i,j),(k,l)\}}(x) = -w_{ij,kl}\beta \log \cosh \left(\frac{x_{ij} - x_{kl}}{\varrho} \right)$$

mit positiven Parametern β und ϱ. Es ist $w_{ij,kl} = 1$ für direkt (orthogonal) benachbarte Pixel (i,j) und (k,l) sowie $w_{ij,kl} = \sqrt{1/2}$ für diagonal benachbarte Pixel[3]. Die Wahl der offenen Parameter β und ϱ wird je nach Untersuchung

[3] In Beispiel 2.23 wurde gezeigt, daß es sich auch bei der allgemeineren, von Green angegebenen Klasse um spezielle Kontrastmodelle handelt.

getroffen und hängt unter anderem von der Körperregion ab. Darüber hinaus können Annahmen des Arztes über die zu erwartenden Konzentrationsverteilungen eine Rolle spielen.

Wichtig für die ausstehende Maximierung der *a-posteriori* Wahrscheinlichkeit ist die Konvexität der sich aus den obigen Potentialen ergebenden Energiefunktion und damit die konkave Form von $\log P(X = x, Y = y)$.

3.4.3 Der EM-Algorithmus

Zur Gewinnung eines MAP-Schätzers für die Rekonstruktion wird wie in Abschnitt 3.3 der Logarithmus der gemeinsamen Verteilung

$$
\begin{aligned}
&\log P(Y = y, X = x) \\
&\quad = \log P(Y = y | X = x) + \log P(X = x) \qquad (3.12) \\
&\quad = \sum_{\varphi \in \Phi} \left(y_\varphi \log \left(\sum_{(i,j) \in L} \gamma_{ij,\varphi} x_{ij} \right) - \sum_{(i,j) \in L} \gamma_{ij,\varphi} x_{ij} - \log(y_\varphi !) \right) - U(x)
\end{aligned}
$$

von Beobachtung und Konzentrationsverteilung maximiert. U bezeichnet die aus den oben erläuterten Cliquenpotentialen zusammengesetzte Energiefunktion der *a-priori* Verteilung. Da y gegeben ist, spielen die Terme $\log(y_\varphi!)$ für die Optimierung keine Rolle, so daß lediglich

$$
\sum_{\varphi \in \Phi} \left(y_\varphi \log \left(\sum_{(i,j) \in L} \gamma_{ij,\varphi} x_{ij} \right) - \sum_{(i,j) \in L} \gamma_{ij,\varphi} x_{ij} \right) - U(x) \qquad (3.13)
$$

zu maximieren ist.

Maximum-Likelihood Rekonstruktion

Betrachtet man zunächst nur das Aufzeichnungsmodell, so wäre die Maximierung der log-Likelihood $\log P(Y = y | X = x)$ erheblich einfacher, wenn statt der Summe der in Detektor φ beobachteten Photonen y_φ die elementareren Anzahlen $z_{ij,\varphi}$ von in Detektor φ aus Pixel (i, j) angekommenen Photonen bekannt wären. Sie sollen der Anschaulichkeit halber *Punktzählungen* heißen. Offensichtlich gilt $y_\varphi = \sum_{(i,j) \in L} z_{ij,\varphi}$, und wegen der Poisson-Charakteristik des radioaktiven Zerfalls folgt für die entsprechende Zufallsvariable $Z_{ij,\varphi}$, daß

$$
Z_{ij,\varphi} \sim \mathcal{POISSON}(\gamma_{ij,\varphi} x_{ij}).
$$

Weil naheliegenderweise auch die $Z_{ij,\varphi}$ stochastisch unabhängig sind, ist es möglich, eine explizite Formel für das Maximum der log-Likelihood $\log P(Z =$

$z|X = x)$ anzugeben. Es gilt nämlich

$$\begin{aligned}
&\arg\max_{x_{ij}} \log P(Z = z|X = x) \\
&\quad = \arg\max_{x_{ij}} \log \prod_{(i,j)\in L} \prod_{\varphi\in\Phi} \left(\frac{(\gamma_{ij,\varphi} x_{ij})^{z_{ij,\varphi}}}{z_{ij,\varphi}!} e^{-\gamma_{ij,\varphi} x_{ij}} \right) \\
&\quad = \arg\max_{x_{ij}} \sum_{(i,j)\in L} \sum_{\varphi\in\Phi} \left(z_{ij,\varphi} \log(\gamma_{ij,\varphi} x_{ij}) - \gamma_{ij,\varphi} x_{ij} \right).
\end{aligned}$$

Für die Maximierung wird der letzte Ausdruck nach x_{ij} abgeleitet

$$\frac{\partial}{\partial x_{ij}} \sum_{(i,j)\in L} \sum_{\varphi\in\Phi} \left(z_{ij,\varphi} \log(\gamma_{ij,\varphi} x_{ij}) - \gamma_{ij,\varphi} x_{ij} \right) = \sum_{\varphi\in\Phi} \left(\frac{z_{ij,\varphi}}{x_{ij}} - \gamma_{ij,\varphi} \right),$$

und man erhält als Nullstelle der Ableitung

$$x_{ij} = \frac{\sum_{\varphi\in\Phi} z_{ij,\varphi}}{\sum_{\varphi\in\Phi} \gamma_{ij,\varphi}}, \tag{3.14}$$

welche das gesuchte Maximum ist. Mit Kenntnis der $z_{ij,\varphi}$ wäre eine Maximierung zumindest der log-Likelihood also unmittelbar durchzuführen.

In der Stochastik bezeichnet man für unmittelbare Problemlösungen zu grobe Beobachtungen als "unvollständige" und den erwünschten Datensatz als "vollständige Daten". Für die Bestimmung von ML- oder MAP-Schätzern aus unvollständigen Daten wurden, zugeschnitten auf spezielle Probleme, von mehreren Autoren iterative Verfahren angegeben. Dempster *et al.* [DLR77] brachten diese in eine einheitliche Form, die unter dem Namen EM-Algorithmus bekannt geworden ist.

Die Buchstaben E und M stehen für die beiden in jeder Iteration durchzuführenden Schritte. Im E-Schritt (*exspectation*) erfolgt die Schätzung der vollständigen Daten unter der Annahme, die aktuelle Lösung des Optimierungsproblems sei bereits die richtige. Mit den so geschätzten Parametern wird im M-Schritt (*maximization*) eine neue Lösung durch Maximierung der Likelihood- bzw. Posterior-Funktion gewonnen. Jede solche Iteration erhöht den Wert der Zielfunktion, und das Verfahren konvergiert schließlich gegen ein (gegebenenfalls nur lokales) Maximum des ML- bzw. MAP-Problems. Eine aktuelle und ausführliche Darstellung, einschließlich der nötigen Beweise, bietet [Tan93].

Im vorliegenden Fall repräsentieren die Zählungen y_φ die unvollständigen Daten, während die Punktzählungen $z_{ij,\varphi}$ vollständige Daten darstellen. Um letztere im E-Schritt zu schätzen, wird der unter der Beobachtung y_φ bedingte

Erwartungswert berechnet, wobei eine (aktuelle) Schätzung x der wahren Aktivität vorauszusetzen ist:

$$\begin{aligned} E([Z_{ij,\varphi}|Y=y]|X=x) &= E([Z_{ij,\varphi}|\sum_{(k,l)\in L} Z_{kl,\varphi}=y_\varphi]|X=x) \\ &\underset{(*)}{=} \frac{y_\varphi E(Z_{ij,\varphi}|X=x)}{E(\sum_{(k,l)\in L} Z_{kl,\varphi}=y_\varphi|X=x)} \\ &= \frac{y_\varphi \gamma_{ij,\varphi} x_{ij}}{\sum\limits_{(k,l)\in L} \gamma_{kl,\varphi} x_{kl}} \end{aligned}$$

Gleichung $(*)$ gilt aufgrund der Faltungsstabilität der Poisson-Verteilung, denn alle $Z_{ij,\varphi}$, $(i,j)\in L$, sind Poisson-verteilt und stochastisch unabhängig[4].

Unter Beschränkung auf das Aufzeichnungsmodell wäre der M-Schritt durch (3.14) gegeben und man erhielte eine ML-Schätzung für das Schnittbild x. Eine Iteration des EM-Algorithmus entstünde aus der Kombination der beiden Schritte und hätte mithin die Form

$$x_{ij}^{(t+1)} = \frac{x_{ij}^{(t)} \sum\limits_{\varphi\in\Phi} \frac{y_\varphi \gamma_{ij,\varphi}}{\sum\limits_{(k,l)\in L} \gamma_{kl,\varphi} x_{kl}^{(t)}}}{\sum\limits_{\varphi\in\Phi} \gamma_{ij,\varphi}}.$$

Die Erfahrungen mit diesem Modell zeigen aber, daß die Ergebnisse stark verrauscht sind, ja mit längerer Laufzeit des Algorithmus sogar noch an Qualität verlieren [Gre90]. Daher wurde von S. Geman und McClure vorgeschlagen, Annahmen über die räumlichen Eigenschaften der Schnittbilder in das Modell zu integrieren [GMcC85, GMcC87]. Anstelle einer ML- wird somit eine MAP-Schätzung bestimmt, in der eine geeignete *a-priori* Verteilung für die Glättung der Rekonstruktion sorgen soll, indem verrauschte Bilder x durch den "Strafterm" $P(X=x)$ geringer gewichtet werden.

[4] Es gilt für stoch. unabh. ZV'en $X \sim \mathcal{POISSON}(\lambda_X)$ und $Y \sim \mathcal{POISSON}(\lambda_Y)$:

$$\begin{aligned} E(X|X+Y=z) &= \sum_{x=0}^{\infty} x \frac{P(X=x, X+Y=z)}{P(X+Y=z)} = \sum_{x=0}^{\infty} x \frac{P(X=x)P(Y=z-x)}{P(X+Y=z)} \\ &= \sum_{x=1}^{z} x \frac{\frac{(\lambda_X)^x}{x!} e^{-\lambda_X} \frac{(\lambda_Y)^{z-x}}{(z-x)!} e^{-\lambda_Y}}{\frac{(\lambda_X+\lambda_Y)^z}{z!} e^{-(\lambda_X+\lambda_Y)}} = \frac{1}{(\lambda_X+\lambda_Y)^z} \sum_{x=1}^{z} \frac{z!\,(\lambda_X)^x (\lambda_Y)^{z-x}}{(x-1)!(z-x)!} \\ &= \frac{z\cdot\lambda_X}{\lambda_X+\lambda_Y} \cdot \frac{1}{(\lambda_X+\lambda_Y)^{z-1}} \sum_{x=0}^{z-1} \binom{z-1}{x} (\lambda_X)^x (\lambda_Y)^{(z-1)-x} \\ &= \frac{z\cdot\lambda_X}{\lambda_X+\lambda_Y} = \frac{z\cdot E(X)}{E(X+Y)} \end{aligned}$$

Der “one step late”-Algorithmus

Bei Hinzunahme einer *a-priori* Verteilung P^X mit Energiefunktion U ist die Zielfunktion (3.13) bezüglich x zu maximieren. Mit den Überlegungen aus dem vorherigen Unterabschnitt reduziert sich das Problem bei Kenntnis der Punktzählungen $z_{ij,\varphi}$ auf die Maximierung von

$$\sum_{(i,j)\in L} \sum_{\varphi\in\Phi} (z_{ij,\varphi} \log(\gamma_{ij,\varphi} x_{ij}) - \gamma_{ij,\varphi} x_{ij}) - U(x).$$

Will man auch hier mit dem EM-Algorithmus vorgehen, so bleibt der E-Schritt unverändert, doch der M-Schritt kann nicht in der oben aufgeführten Weise vollzogen werden, weil wegen der komplizierten Form von $\frac{\partial}{\partial x_{ij}} U(x)$ das zu (3.14) analoge Maximum nicht ohne weiteres berechenbar ist.

Green schlägt daher vor, anstelle der allgemeinen Energiefunktion ihren Wert an der zuletzt als Lösung vermuteten Stelle $x^{(t)}$ zu verwenden und bezeichnet diese Variante des EM-Algorithmus als *one step late*-Algorithmus (OSL). Analog zu oben erhält man dann für den M-Schritt der $(t+1)$-ten Iteration als Maximum an der Stelle (i,j)

$$\frac{\sum_{\varphi\in\Phi} z_{ij,\varphi}}{\sum_{\varphi\in\Phi} \gamma_{ij,\varphi} + \frac{\partial}{\partial x_{ij}} U(x)\Big|_{x=x^{(t)}}}.$$

Offensichtlich wird dadurch in jeder Iteration das wahre Maximum des M-Schritts nur angenähert, aber empirische Ergebnisse [Gre90] mit den erwähnten Potentialen deuten darauf hin, daß weiterhin Konvergenz vorliegt. Des weiteren gibt Green eine verbesserte Version des OSL-Algorithmus an, die hier nicht von Interesse ist, für die es Lange [Lan90] aber gelang, die Konvergenz unter einer sehr allgemeinen Klasse von *a-priori* Verteilungen nachzuweisen.

3.4.4 Beispiel und Anmerkungen

Das Beispiel in Abbildung 3.21 stammt aus [Gre90] und zeigt ein horizontales Schnittbild des Patientenkopfes. Man erkennt deutlich die Ansammlung des Radiopharmazeutikums in den Seitenhörnern des Liquorraums, dem Stoßdämpfer und Nahrungslieferanten des Gehirns.

Für die Rekonstruktion wurden 16 Iterationen des OSL-Algorithmus, beginnend mit einem Bild konstanten Grauwerts, durchgeführt. Anschließend wurden die Gewichte $\gamma_{ij,\varphi}$ angepaßt und der Algorithmus lief für weitere 128 Iterationen, nach denen keine sichtbaren Änderungen mehr auftraten. Die

Abbildung 3.21: OSL-Rekonstruktion nach 16+128 Iterationen

Anpassung der Gewichte und die konkrete, verbesserte Implementation des OSL-Algorithmus sind in [Gre90] beschrieben.

Die geringe Auflösung des Bildes ist typisch für die E-CT. Ihre Berechtigung gegenüber anderen, höher auflösenden Verfahren wie Transmissions-CT oder Magnetic-Resonance-Tomographie (MR) erlangt die E-CT durch den andersgearteten Anwendungszweck. Statt der vorwiegend anatomischen Informationen der anderen Methoden lassen sich mit der E-CT Stoffwechseleigenschaften visualisieren.

Den ersten praktikablen Ansatz zur Bildrekonstruktion mit Hilfe stochastischer Modellierung beschrieben Shepp und Vardi [SV82]. Sie berechneten eine ML-Schätzung für die Rekonstruktion von Positronen-Emissions-Tomographenbildern (PET). Bei der Kernumwandlung durch Positronenzerfall wandelt sich ein Proton des Atomkerns in ein Neutron um, wobei ein positiv geladenes Elektron ("Positron") ausgeschleudert wird. Nach Abgabe seiner kinetischen Energie zerstrahlt es zusammen mit einem nahen Elektron, und vom Vereinigungspunkt werden zwei Photonen in entgegengesetzte Richtungen ausgesandt. Die Behandlung der Absorptionseffekte vereinfacht sich nicht nur durch die höhere Strahlungsenergie von 511 keV, sondern auch, weil sich die Photonen jedes dieser Paare zusammengenommen entlang einer kompletten Körpertransversale bewegt hat, deren Absorption sich durch eine Vormessung bestimmen läßt. Andererseits ist SPECT das technisch günstigere und weiter verbreitete Verfahren, weil die für PET benötigten Radionuklide (wie etwa ^{11}C, ^{13}N, ^{15}O oder ^{18}F) sehr kurze Halbwertszeiten im Bereich von 2–110 Minuten (SPECT: 6 Stunden bis einige Tage) haben und daher am Einsatzort erzeugt werden müssen [Win90, Gre90].

Die Glättung der recht "rauhen" ML-Rekonstruktionen durch die lokalen Abhängigkeiten einer *a-priori* Gibbs-Verteilung geht wie erwähnt auf S. Geman und McClure [GMcC85, GMcC87] zurück, die sich vom Bayes-Ansatz insgesamt eine Vereinheitlichung der Behandlung zahlreicher Probleme der Bildverarbeitung versprechen [GMcC85].

Lipinski [Lip94] untersucht derzeit die Verwendung von MR-Bildern zur Verbesserung der *a-priori* Verteilung. Aufgrund der MR-Information wird das Nachbarschaftssystem des Schnittbildes modifiziert, indem Bindungen zwischen Pixeln unterschiedlicher anatomischer Regionen aufgehoben werden. Im wesentlichen entspricht dies der Schätzung von Kantenelementen (vgl. Abschnitt 3.2) aus zusätzlicher Information. Es spiegelt sich darin die Tatsache wider, daß sich das Radiopharmazeutikum in verschiedenen anatomischen Regionen unterschiedlich stark konzentriert, dafür aber innerhalb der Regionen sehr gleichmäßig.

3.5 Ergänzungen

Der Trend zur Modellierung der Klasse von Bildern, auf die ein Verfahren angewandt wird, hat sich in den 80er Jahre zügig verbreitet und durchzieht mittlerweile fast alle Bereiche der Bildverarbeitung [ChJ93, Kapitel 1]. Um dies zu unterstreichen, werden in diesem Abschnitt einige weitere Anwendungsgebiete aufgeführt, in denen Markoff-Felder zum Einsatz kommen.

Zuvor jedoch wird noch kurz auf ein weiteres deterministisches Verfahren zur Schätzerbestimmung hingewiesen und darauf eingegangen, inwieweit die vorgestellten Lösungsansätze einem einheitlichen Schema folgen.

3.5.1 Iterated Conditional Modes

Die in den vorangegangenen Abschnitten dargestellten Problemlösungen erforderten jeweils die Berechnung eines Schätzers. Die dazu verwendeten Verfahren waren Simulated Annealing, Dynamische Programmierung und der EM-Algorithmus. Ein wichtiges weil einfaches Verfahren blieb dabei unberücksichtigt und soll an dieser Stelle noch erwähnt werden.

Besag [Bes86] schlägt vor, zur Bestimmung eines MAP-Schätzers iterativ vorzugehen und in jedem Schritt die unter der Beobachtung und den aktuellen Werten der übrigen Komponenten bedingte Wahrscheinlichkeit einer Komponente, also

$$P(X_{ij} = x_{ij} | Y = y, X_{L \setminus \{(i,j)\}} = x_{L \setminus \{(i,j)\}}),$$

bezüglich x_{ij} zu maximieren. Konkret geschieht dies durch die dazu äquivalente Maximierung von

$$P(Y = y | X = x) \cdot P(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}})$$

bezüglich x_{ij}, wobei sich der zu berücksichtigende Teil der Likelihood $P(Y = y | X = x)$ abhängig vom Fehlermodell noch vereinfacht. Besag nennt diesen Algorithmus *Iterated Conditional Modes (ICM)* und verwendet ihn in [Bes86] zur Bildrestauration. Wegen

$$\begin{aligned} & P(X = x | Y = y) \\ &= P(X_{ij} = x_{ij} | Y = y, X_{L \setminus \{(i,j)\}} = x_{L \setminus \{(i,j)\}}) \\ &\quad \cdot P(X_{L \setminus \{(i,j)\}} = x_{L \setminus \{(i,j)\}} | Y = y) \end{aligned}$$

wird die *a-posteriori* Wahrscheinlichkeit bei jeder Iteration höchstens vergrößert, so daß ICM gegen ein lokales Maximum konvergiert. Als Startbild bietet sich daher die häufig mit wenig Aufwand zu berechnende ML-Schätzung an, also jenes $x \in \Gamma^L$, für das $P(Y = y | X = x)$ maximal wird. Es scheint, daß Konvergenz dann in der Regel schon nach wenigen Durchläufen aller Pixel vorliegt [Bes86].

Schnelle Konvergenz und einfache Implementierung sind die Hauptgründe dafür, daß ICM in Anwendungen, in denen es mehr auf Geschwindigkeit denn auf möglicherweise höhere Qualität ankommt, ein gern benutztes Verfahren ist [A. Kaup[5], mündliche Mitteilung]. So setzen Aach *et al.* [AKM93] ICM zur Detektion von Bewegung in Bildsequenzen ein. Sie merken überdies an, daß die lediglich lokale Konvergenz kein wesentlicher Nachteil sei, weil das Ausgangsbild, auf das ICM zur Glättung angesetzt wird, gut genug sei, um zu einem akzeptablen Resultat zu führen.

Bei genauerer Betrachtung fällt allerdings auf, daß ICM im Grunde doch schon behandelt wurde. Werden nämlich vollständige Daten y beobachtet, so entspricht ICM dem Spezialfall des EM-Algorithmus aus Abschnitt 3.4 unter Auslassung des dann überflüssigen E-Schrittes. Wenn in Abschnitt 4.2 der Gibbs-Sampler eingeführt sein wird, wird weiterhin klar werden, daß diese Einschränkung des EM-Algorithmus' wiederum einem Gibbs-Sampler mit sofortigem "Einfrieren" (Temperaturparameter 0) gleichkommt.

[5]Institut für Elektrische Nachrichtentechnik, RWTH Aachen.

3.5.2 Schematisierung von Problemlösungen

Zu den interessantesten Aspekten der stochastischen Bildmodellierung gehört die Beobachtung, daß die daraus enstehenden Lösungsansätze einem einheitlichen Schema folgen.

In [GMcC85] geben S. Geman und McClure der Hoffnung Ausdruck, durch das Bayes-Schema eine Vereinheitlichung weiter Bereiche der Bildverarbeitung zu erzielen. Die hier vorgestellten Ansätze haben gezeigt, daß Anlaß zu dieser Hoffnung besteht, denn die in ihnen vorgeschlagenen Lösungen bestanden trotz der Verschiedenheit der Probleme jeweils in der Bestimmung eines MAP-Schätzers, der durch die Bayes-Formel in die gemeinsame Maximierung von Likelihood und *a-priori* Verteilung umformuliert wurde. Man beachte dabei, daß die Maximierung der logarithmierten gemeinsamen Verteilungen (3.8) und (3.12) gerade der Minimierung der *a-posteriori* Energiefunktion (3.6) entspricht.

Abgesehen von der in diesem Kontext weniger verbreiteten Dynamischen Programmierung sind die verwendeten Berechnugsverfahren verschiedene Ausprägungen derselben Vorgehensweise, die in Kapitel 4 formalisiert wird. Zusammengenommen ergibt sich damit etwa folgendes allgemeine Schema (siehe auch Abbildung 3.22):

1.	Modellierung des Zielbildes und möglichst vieler der darin angenommenen lokalen Abhängigkeiten durch ein MRF X
2.	Aufstellung eines Abbildungs- bzw. Fehlermodells
3.	Modellierung der aus X aufgrund des Abbildungs- bzw. Fehlermodells möglichen Beobachtungen durch ein MRF Y
4.	Maximierung von $P(X = x \mid Y = y)$ durch Maximierung von $P(Y = y \mid X = x) \cdot P(X = x)$ bezüglich x für das beobachtete y

Für Schritt 4 bietet sich grundsätzlich Simulated Annealing an, das zur Beschleunigung gegebenfalls auf ICM beschränkt werden kann. Für unvollständige Beobachtungen y sollte das Verfahren um den Schätzschritt zum EM-Algorithmus erweitert werden. Bei günstiger Form der *a-posteriori* Verteilung sind aber auch spezielle Optimierungsverfahren denkbar (wie etwa in Abschnitt 3.4 die Dynamische Programmierung).

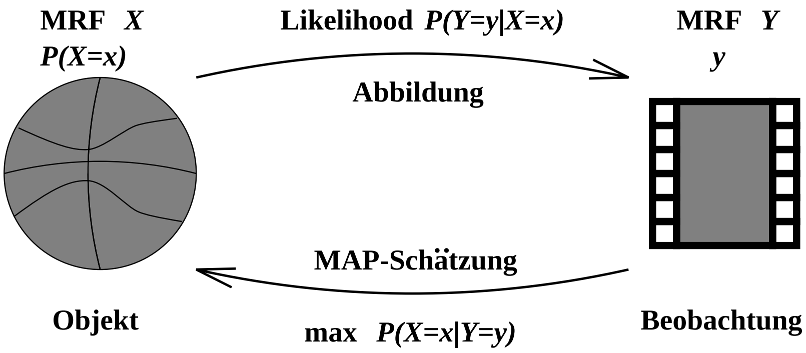


Abbildung 3.22: Schematisierte Modellbildung

3.5.3 Weitere Anwendungen

Eine Auswahl von Anwendungsfeldern soll unterstreichen, in welchen anderen Bereichen das Markoff-Modell nützlich sein kann. Es wird kein Anspruch auf Vollständigkeit erhoben und auch nicht näher auf die Methoden eingegangen.

Neben den in den vorausgegangenen Abschnitten bereits aufgeführten Anwendungen haben Markoff-/Gibbs-Felder und Varianten davon auf den folgenden Gebieten zu neuartigen Ansätzen geführt[6]:

- Kontur-Erkennung
- Codierung und Kompression von Bildern und Bildsequenzen
- 2D- und 3D-Klassifikation
- Computergrafik, insbesondere 2D- und 3D-Texturierung
- Stereobild-Analyse
- Bewegungsanalyse
- Early Vision
- Bildinterpretation
- Halftoning

Durch die Entlehnung aus der statistischen Physik liegt auf der Hand, daß das Modell nicht nur in der Bildverarbeitung Anwendung findet (worauf in Kapitel 5 kurz eingegangen wird). Interessant daran ist, daß viele der in anderen Bereichen entwickelten Konzepte auf die Bildverarbeitung übertragbar

[6]Die Liste ist das Ergebnis der Auswertung einer INSPEC-Literaturrecherche.

sind. Beispielsweise ist das *multiresolution*-Modell zur Effizienzsteigerung von MRF-Algorithmen inspiriert durch das physikalische Modell der Renormierungsgruppen [ChJ93, Kapitel 6].

Kapitel 4

Sampling und Optimierung

> VALERIO: Aber eigentlich wollte ich einer hohen und geehrten Gesellschaft verkündigen, daß hiemit die zwei weltberühmten Automaten angekommen sind und daß ich vielleicht der dritte und merkwürdigste von beiden bin, wenn ich eigentlich selbst recht wüßte, wer ich wäre, worüber man übrigens sich nicht wundern dürfte, da ich selbst gar nichts von dem weiß, was ich rede, ja auch nicht einmal weiß, daß ich es nicht weiß, so daß es höchst wahrscheinlich ist, daß man mich nur so reden läßt, und es eigentlich nichts als Walzen und Windschläuche sind, die das Alles sagen.
>
> G. BÜCHNER: LEONCE UND LENA

In diesem Kapitel soll die bisher unbeantwortete Frage behandelt werden, wie aus der gegebenen Verteilung eines Gibbs-Feldes Bildrealisierungen gewonnen werden. Dieser Vorgang wird meist als "Ziehen" (*sampling*) aus der Verteilung bezeichnet. Ferner wird z.B. für die Bestimmung eines MAP-Schätzers ein unter einer *a-posteriori* Verteilung maximal wahrscheinliches Bild (*mode*) benötigt. Auch dieser Fall wird behandelt.

Für die formale Beschreibung der Vorgehensweise wird ein allgemeines Maschinenmodell eingeführt[1], das im wesentlichen eine Erweiterung der aus der Klassifikation und kombinatorischen Optimierung bekannten Boltzmann-Maschine (BM) [AK89] darstellt. Für diese Erweiterung, die *Gibbs-Maschine* genannt werden soll, werden zwei Berechnungsvorschriften angegeben, welche wiederum Varianten des Simulated Annealing (SA) sind. Es handelt sich dabei zum einen um den von Geman & Geman [GG84] vorgeschlagenen Gibbs-Sampler (vgl. auch Abschnitt 3.2), und zum anderen um die erste und be-

[1] Der Entwurf des Maschinenmodells erfolgte noch in Unkenntnis des Modells von Azencott [Aze90], das in einigen Punkten Ähnlichkeiten aufweist.

kannteste Form des SA, nämlich den Metropolis-Algorithmus [MRT53]. Speziell der Gibbs-Sampler kann als Verallgemeinerung des Berechnungsschemas der Boltzmann-Maschine angesehen werden.

Das Maschinenmodell läßt sich leicht zu einem synchronen, (eingeschränkt) parallelen Modell erweitern. Leider ist jedoch noch nicht bekannt, ob auch eine asynchron parallele Arbeitsweise die gewünschten Resultate liefert. In Abschnitt 4.4 wird kurz auf die entstehenden Probleme eingegangen.

Schließlich wird noch gezeigt, daß auch die Bestimmung einer MPL-Schätzung der Parameter eines Gibbs-Feldes (siehe Abschnitt 3.1.3) durch einen der Gibbs-Maschine analogen Formalismus beschreibbar ist.

4.1 Die Gibbs-Maschine

Es soll zunächst eine formale Beschreibung dafür gefunden werden, daß mittels eines weitgehend beliebigen Algorithmus' gemäß der zugehörigen Gibbs-Verteilung die Realisierung eines Markoff-Zufallsfeldes erzeugt wird. Zu diesem Zweck wird der Ablauf des Algorithmus durch eine konvergente Markoff-Kette beschrieben, deren stationäre Grenzverteilung gerade der gewünschten Gibbs-Verteilung entspricht.

Ausgehend von einem beliebigen Startbild $x^{(0)}$ wird zu jedem Zeitpunkt $t \in \mathbb{N}_0$ ein Pixel ausgewählt, das gemäß einer noch zu bestimmenden Update-Regel einen neuen Grauwert erhält. Die Ersetzung erfolgt mit dem Ziel, die Energie des Bildes immer weiter zu verringern, seine Wahrscheinlichkeit als Realisierung der Gibbs-Verteilung also zu erhöhen.

Es ist klar, daß nicht jede Update-Regel dies leisten wird. Weil aber zumindest zwei verschiedene Regeln gebräuchlich sind und vorgestellt werden sollen, erfolgt die Definition der Gibbs-Maschine unabhängig von der später einzusetzenden Update-Regel. Der Einfachheit der Notation wegen werden hier nur Gibbs-Maschinen für "reine" Zufallsfelder, also solche mit einheitlicher Grauwertmenge für alle Komponenten, behandelt. Die Erweiterung auf "gemischte" Zufallsfelder (wie beispielsweise die um Kantenelemente erweiterten Felder aus Abschnitt 3.2) ist problemlos.

Definition 4.1 (Gibbs-Maschine)
Es sei X ein Gibbs-Feld über L bezüglich η mit Potential V und Temperatur T. Dann heißt $\mathcal{G} = (\Gamma, L, \eta, V, s, c, \Upsilon)$ (sequentielle) Gibbs-Maschine (GM) *zu X, falls*

- *$s : \mathbb{N}_0 \to L$ eine Funktion mit $L \subseteq \{s(m), \ldots, s(m+n-1)\}$ für ein $n \in \mathbb{N}$ und alle $m \in \mathbb{N}_0$ ist,*

- $c : \mathbb{N}_0 \to \mathbb{R}^+$ *eine monoton fallende Funktion mit* $\lim_{t\to\infty} c(t) = T$ *ist und*
- $\Upsilon^{(t)} : \Gamma^{\overline{\eta_{s(t)}}} \to \{X^{(t)}_{s(t)} | X^{(t)}_{s(t)} : (\Omega, \mathcal{A}) \to (\Gamma, \mathcal{P}(\Gamma))\}$ *für jedes* $t \in \mathbb{N}_0$ *eine Funktion ist, die Nachbarschaftskonfigurationen auf Grauwertzufallsvariablen abbildet.*

Dabei heißen s Scan-Funktion, *c* Abkühlschema (cooling-schedule) *und* $\Upsilon = \{\Upsilon^{(t)}\}_{t\in\mathbb{N}_0}$ Update-Regel *der GM.*

Wie angedeutet besteht der Zweck der Gibbs-Maschine darin, auf die Eingabe eines beliebigen Bildes $x^{(0)}$ eine bestimmte Realisierung x des Zufallsfeldes X mit Wahrscheinlichkeit $P(X = x)$ zu liefern. Die Definition der GM ist so angelegt, daß sie das Gerüst eines Algorithmus' für dieses Problem widerspiegelt. Die Wahl der drei noch offenen Komponenten bestimmt die konkrete Implementation, von der hier abstrahiert werden sollte. Insbesondere kann die Update-Regel als austauschbares Programm-Modul angesehen werden. Abstrakt liefert sie zu jedem Berechnungsschritt eine Zufallsvariable, welche die Wahrscheinlichkeit der möglichen neuen Grauwerte für den von der Scan-Funktion bestimmten Bildpunkt angibt. Da diese, unter Rückgriff auf die Markoff-Eigenschaft (2.3), nur von den Grauwerten der Nachbarschaft abhängen soll, besteht der Definitionsbereich ausschließlich aus den möglichen aktuellen Nachbarschaftskonfigurationen $\Gamma^{\overline{\eta_{s(t)}}}$.

Aus dem letzten Argument ergibt sich – unabhängig von der verwendeten Update-Regel – sofort das Parallelpotential jeder Gibbs-Maschine, da die den Ersetzungsschritt beschreibende Zufallsvariable eben *allein* von $x^{(t)}_{\overline{\eta_{s(t)}}}$ abhängig gemacht werden kann. Andererseits ist auf diese Weise zunächst nur gesichert, daß Bildpunkte mit *disjunkten* geschlossenen Nachbarschaften parallel aktualisiert werden können. Abschnitt 4.4 wird sich kurz mit den auftretenden Fragen und Problemen befassen.

Beispiel 4.2 (GM für Kontrast-Modelle)
Wie bei allen Auto-Modellen der Fall, können Gibbs-Maschinen zu δ*-Kontrast-Modellen als Graph mit Kontenmenge L repräsentiert werden. Jeder Knoten erhält einen Grauwert als Label. Zwei Knoten (Bildpunkte)* (i, j) *und* (k, l) *sind durch eine Kante verbunden, wenn sie in* η *benachbart sind. Die Kante erhält die Bindungsstärke* $\beta_{ij,kl}$ *als Gewicht. Abbildung 4.1 veranschaulicht diese Konstruktion für ein kleinen Bildausschnitt. Der durch die GM formalisierte Algorithmus wählt dann durch die Scan-Funktion s zu jedem Zeitpunkt einen Knoten aus und kann allein aufgrund der Nachbarn dieses Knotens und der globalen Parameter* α_g *über das neue Grauwertlabel entscheiden. Vorzugsweise verfährt s entlang eines Hamiltonkreises durch den Graphen.*

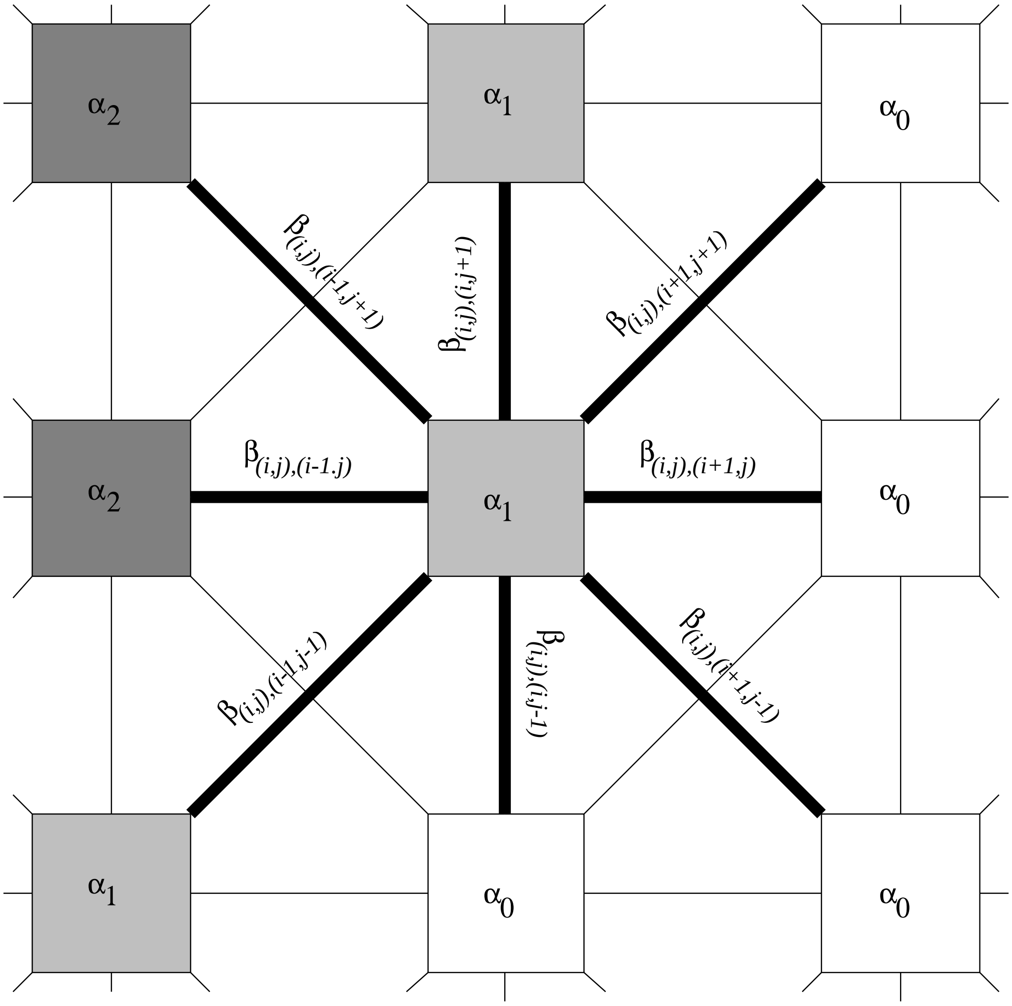


Abbildung 4.1: Graph-Repräsentation der GM zu einem δ-Kontrast-Modell. Die Knoten sind entsprechend dem aktuellen Grauwert des zugehörigen Pixels eingefärbt und mit dem korrespondierenden globalen Parameter markiert, die Kanten mit den Bindungsstärken $\beta_{ij,kl}$ gewichtet.

Um nun die Berechnung eines Zielbildes x formal zu fassen, wird weiterhin definiert:

Definition 4.3 ((Folge-)Konfiguration, Berechnung)
Die Menge der Konfigurationen *einer Gibbs-Maschine $\mathcal{G}$ ist gegeben durch $\Gamma^L \times I\!N_0$. Eine Konfiguration $(y, t+1)$ heißt* Folgekonfiguration *von (x, t), in Zeichen*

$$(x, t) \vdash_{\mathcal{G}} (y, t+1),$$

falls $y = x^{y_{s(t)}}$ und $P(\Upsilon^{(t)}(x_{\overline{\eta_{s(t)}}}) = y_{s(t)}) > 0$. Eine Folge $\{(x^{(t)}, t)\}_{t \in I\!N_0}$ heißt Berechnung *der Gibbs-Maschine $\mathcal{G}$, falls $(x^{(t)}, t) \vdash_{\mathcal{G}} (x^{(t+1)}, t+1)$ für alle $t \in I\!N_0$.*

Die möglichen Berechnungen einer Gibbs-Maschine sind dann beschrieben durch eine Markoff-Kette $\{X^{(t)}\}_{t \in I\!N_0}$ von Zufallsfeldern mit Übergangswahrscheinlichkeiten

$$P(X^{(t+1)} = y | X^{(t)} = x) = \begin{cases} P\left(\Upsilon^{(t)}(x_{\overline{\eta_{s(t)}}}) = y_{s(t)}\right) & \text{falls } y = x^{y_{s(t)}} \\ 0 & \text{sonst,} \end{cases} \tag{4.1}$$

wobei $P^{X^{(0)}}$ die Verteilung der Startbilder ist.

Eine Update-Regel ist natürlich nur dann sinnvoll, wenn sie zusammen mit Abkühlschema und Scan-Funktion die Konvergenz der obigen Markoff-Kette gegen die Verteilung eines Markoff- bzw. Gibbs-Feldes X sichert. Eine GM mit einer in dieser Hinsicht sinnvollen Update-Regel soll konvergent genannt werden:

Definition 4.4 (Konvergenz einer Gibbs-Maschine)
Eine Gibbs-Maschine $\mathcal{G}$ zu einem Markoff-Feld X heißt konvergent, *wenn mit den obigen Bezeichnungen bei beliebiger Startverteilung $P^{X^{(0)}}$ für jedes Bild $x \in \Gamma^L$ gilt:*

$$\lim_{t \to \infty} P(X^{(t)} = x) = P(X = x) \quad \left[= \frac{1}{Z(T)}\, e^{-U(x)/T} \right]$$

4.2 Spezielle Update-Regeln

Definition 4.4 konkretisiert nun zwar, wann eine GM das Gewünschte leistet, die *Existenz* einer Update-Regel, welche die Konvergenz der GM garantiert, ist aber noch zu zeigen.

In diesem Abschnitt werden daher die beiden im Umgang mit Markoff-Feldern gebräuchlichsten Verfahren, beide Varianten des Simulated Annealing, als Update-Regeln in den vorgestellten Formalismus eingebettet.

4.2.1 Updating mit dem Metropolis-Algorithmus

Simulated Annealing (SA) ist ein stochastisches Verfahren zur approximativen Optimierung. Seine bekannteste Form wurde bereits in den 50er Jahren von Metropolis *et al.* [MRT53] verwendet, um Erwartungswerte für Zustandsgrößen von Materialkonfigurationen (z.B. bei Kristallen) zu berechnen. Die Idee des Metropolis-Algorithmus besteht darin, von einer beliebigen Startkonfiguration ausgehend kleine Änderungen daran vorzuschlagen und diese immer dann zu akzeptieren, wenn sie eine Verbesserung der Zielfunktion bewirken. Um nicht in lokalen Optima steckenzubleiben, werden mit gewissen, vom Grad der Verschlechterung abhängigen Wahrscheinlichkeiten auch nachteilige Änderungsvorschläge angenommen. Für die Konvergenz ist es allerdings nötig, daß die Akzeptanz von Verschlechterungen mit Fortschreiten des Algorithmus' abnimmt. Eine umfassende Darstellung dieses und anderer, verwandter Verfahren findet sich bei Hammersley und Handscomb [HH64].

Für die Gibbs-Maschine wird nun eine Update-Regel angegeben, die den Bedingungen des Metropolis-Algorithmus' genügt und fortan auch als solcher bezeichnet wird. Zuvor sei noch der *Positivteil* a^+ einer reellen Zahl a definiert durch

$$a^+ := \begin{cases} a & \text{für } a > 0 \\ 0 & \text{sonst.} \end{cases}$$

Definition 4.5 (Metropolis-Algorithmus)
Es sei $\mathcal{G} = (\Gamma, L, \eta, V, s, c, \Upsilon)$ eine Gibbs-Maschine. Für $y_{s(t)} \in \Gamma$ heißt dann die durch

$$P\left(\Upsilon^{(t)}(x_{\overline{\eta_{s(t)}}}) = y_{s(t)}\right) \tag{4.2}$$
$$= \begin{cases} \frac{1}{G}\exp\left\{-\frac{(U(x^{y_{s(t)}})-U(x))^+}{c(t)}\right\} & \textit{falls } y_{s(t)} \neq x_{s(t)} \\ 1 - \sum\limits_{\substack{y_{s(t)}\in\Gamma \\ y_{s(t)}\neq x_{s(t)}}} \frac{1}{G}\exp\left\{-\frac{(U(x^{y_{s(t)}})-U(x))^+}{c(t)}\right\} & \textit{sonst} \end{cases}$$

gegebene Update-Regel Metropolis-Algorithmus (MA).

Formel (4.2) drückt aus, daß zum Zeitpunkt $t \in \mathbb{N}_0$ der Ersetzungsschritt des Grauwertes am Bildpunkt $s(t)$ in zwei stochastisch unabhängige Teile zerfällt. Der Vorschlag für den Folgezustand wird gleichverteilt auf der Menge der möglichen Grauwerte ermittelt[2], woraus sich der erste Faktor der rechten Seite ergibt. Ist das so generierte $y_{s(t)}$ ungleich $x_{s(t)}$, wird es mit der

[2] Die Gleichverteilung ist nicht zwingend [AK89], wegen ihrer Einfachheit aber die gebräuchlichste Generierungsgrundlage.

durch den zweiten Faktor der rechten Seite gegebenen Wahrscheinlichkeit akzeptiert. Tritt eine Verminderung der Energie ein, so ist dieser Faktor 1, ansonsten hängt er vom Grad der Verschlechterung und der aktuellen Temperatur $c(t)$ ab. Die Wahrscheinlichkeit dafür, daß die Komponente mit Index $s(t)$ unverändert bleibt, ergibt sich aus dem Komplement der Änderungswahrscheinlichkeiten.

Trotz der auftretenden globalen Energiedifferenz ist die Verteilung der Zufallsvariablen $\Upsilon^{(t)}(x_{\overline{\eta_{s(t)}}})$ wohldefiniert. Aufgrund von (2.5) gilt nämlich

$$\begin{aligned} U(x^{y_{s(t)}}) - U(x) &= \sum_{C \in \mathcal{C}} V_C(x^{y_{s(t)}}) - V_C(x) \\ &= \sum_{C \in \mathcal{C} : s(t) \in C} V_C(x^{y_{s(t)}}) - V_C(x), \end{aligned}$$

d.h. die Differenz hängt tatsächlich nur von den Punkten in $\overline{\eta_{s(t)}}$ ab.

Das folgende Lemma rechtfertigt die Verwendung von MA als Update-Regel für Gibbs-Maschinen.

Lemma 4.6 (Konvergenz von MA)
Es sei $\mathcal{G}$ eine Gibbs-Maschine zu einem Gibbs-Feld X mit Update-Regel MA. Dann ist $\mathcal{G}$ konvergent für jedes $T > 0$, unabhängig vom Abkühlschema c. Weiterhin ist $\mathcal{G}$ konvergent für den Übergang $T \to 0$, falls

$$c(t) = \frac{\Delta \cdot |L|}{\log(\left\lfloor \frac{t}{|L|} \right\rfloor + 1)}$$

mit

$$\Delta \geq \max_{x,y \in \Gamma^L} \{U(y) - U(x)\}. \tag{4.3}$$

Da sich die in [AK89] geführten Beweise für den allgemeinen Metropolis-Algorithmus auf den vorliegenden Fall übertragen lassen, wird von einem Nachweis des Lemmas abgesehen. Man beachte jedoch, daß der Fall $T > 0$ dem Konvergenzbeweis für die homogenen Markoff-Ketten (ohne Annealing) entspricht und daher insbesondere das konstante Abkühlschema $c(t) \equiv T$ zugelassen ist.

4.2.2 Updating mit dem Gibbs-Sampler

Wie in Abschnitt 3.2 angekündigt, wird an dieser Stelle der Mechanismus erläutert, den Geman & Geman vorschlugen, um das Maximum der *a-posteriori* Verteilung gestörter Bilder zu approximieren [GG84].

Die dahinterstehende Idee weicht von der des Metropolis-Algorithmus insofern ab, als daß nicht gleich- oder sonstwieverteilt ein Nachfolgezustand generiert und dieser dann entsprechend seinem Nutzen für die anvisierte Verteilung gewichtet wird, sondern daß er unmittelbar aus dieser gezogen wird (wobei die durch das Abkühlschema gegebene Temperatur zu berücksichtigen ist).

In der Update-Regel manifestiert sich das Ziehen aus einer Verteilung, die der des Gibbs-Feldes X bei Temperatur $c(t)$ entspricht, in der Wahl der sich bei dieser Temperatur ergebenden lokalen Charakteristiken von x als Verteilung der Zufallsvariablen für den neuen Grauwert.

Definition 4.7 (Gibbs-Sampler)
Es sei $\mathcal{G} = (\Gamma, L, \eta, V, s, c, \Upsilon)$ *eine Gibbs-Maschine. Für* $y = x^{y_{s(t)}}$ *heißt die durch*

$$P\left(\Upsilon^{(t)}(x_{\overline{\eta_{s(t)}}}) = y_{s(t)}\right) = \frac{1}{Z_{s(t)}(c(t))} \exp\left\{-\frac{\sum\limits_{C \in \mathcal{C}: s(t) \in C} V_C(y)}{c(t)}\right\} \tag{4.4}$$

mit

$$Z_{s(t)}(c(t)) = \sum_{y_{s(t)} \in \Gamma} \exp\left\{-\frac{\sum\limits_{C \in \mathcal{C}: s(t) \in C} V_C(x^{y_{s(t)}})}{c(t)}\right\}$$

gegebene Update-Regel Gibbs-Sampler (GS).

Ähnlich wie beim Metropolis-Algorithmus zeigt eine Analyse der durch die möglichen Berechnungen bestimmten Markoff-Kette $\{X^{(t)}\}_{t \in \mathbb{N}_0}$ die Konvergenz der Gibbs-Maschine mit Update-Regel GS.

Lemma 4.8 (Konvergenz von GS)
Es sei $\mathcal{G}$ *eine Gibbs-Maschine zu einem Gibbs-Feld* X *mit Update-Regel GS. Dann ist* $\mathcal{G}$ *konvergent für jedes* $T > 0$, *unabhängig vom Abkühlschema c. Weiterhin ist* $\mathcal{G}$ *konvergent für den Übergang* $T \to 0$, *falls*

$$c(t) = \frac{\Delta \cdot |L|}{\log(\left\lfloor \frac{t}{|L|} \right\rfloor + 1)}$$

mit

$$\Delta \geq \max_{x,y \in \Gamma^L} \{U(y) - U(x)\}.$$

Wiederum wird von dem länglichen Beweis des Lemmas abgesehen, weil der entsprechende Beweis aus [GG84] ohne größere Schwierigkeiten in den generelleren Rahmen übertragbar ist.

Außerdem sieht man nun leicht, daß jede Boltzmann-Maschine einer Gibbs-Maschine mit Gitter und Nachbarschaftssystem gemäß der Topologie der BM und Grauwertmenge $\Gamma = \{0, 1\}$ entspricht. Dazu weist man Einer- und Zweiercliquen, deren Punkte alle den Wert 1 haben, das Kantengewicht der BM als Cliquenpotential zu, alle anderen Cliquenpotentiale werden auf Null gesetzt. Schließlich betrachtet man mit GS als Update-Regel den Übergang $T \to 0$. Um für $y_{ij} = 1 - x_{ij}$ die Korrespondenz von GS-Updating und der Umpolwahrscheinlichkeit

$$\frac{1}{1 + e^{(U(x^{y_{ij}}) - U(x))/c(t)}}$$

der BM [Ruf93] einzusehen, erweitere man letztere lediglich um $e^{-U(x^{y_{ij}})/c(t)}$ und erhalte nach einfacher Umformung (4.4).

4.2.3 Vergleich der Update-Regeln

Die beiden oben angeführten Update-Regeln folgen zwei völlig verschiedenen Paradigmen. Beim Metropolis-Algorithmus wird eine symmetrische, irreduzible, aperiodische Markoff-Kette, gegeben durch die Gleichverteilung für die Auswahl möglicher Nachfolgegrauwerte, mit Akzeptanzwahrscheinlichkeiten gewichtet. Der Gibbs-Sampler hingegen zieht die Nachfolgekonfigurationen unmittelbar aus der gewünschten Zielverteilung. In beiden Fällen jedoch konvergieren die Verteilungen der Markoff-Ketten gegen die Verteilung des zur Gibbs-Maschine gehörigen Zufallsfeldes (bei der durch den Grenzwert des Abkühlschemas gegebenen Temperatur).

Die Entscheidung für eine der beiden Update-Regeln hängt vornehmlich davon ab, wie schnell ein Update unter den konkreten Umständen berechnet werden kann, da die Konvergenzgeschwindigkeit auf Grund der identischen Abkühlschemata aus Lemma 4.6 bzw. 4.8 theoretisch bedeutungslos ist.

Ein prinzipieller Vorteil des MA liegt in der Vermeidung der häufig zu komplexen Zustandssummen $Z_{s(t)}(c(t))$. Andererseits kann sich bei homogenen Markoff-Feldern mit nicht zu großer Grauwertmenge die Vorabberechnung der Zustandssummen aller Nachbarschaftskonfigurationen jeweils zu Beginn eines konstanten Intervalls des Abkühlschemas lohnen. Sind diese homogenen Teilstücke der Markoff-Kette $\{X^{(t)}\}_{t \in \mathbb{N}_0}$ lang genug, scheint GS die bessere Wahl zu sein. Insbesondere gilt dies damit für das Sampling bei konstantem Abkühlschema c.

4.3 Einsatzweise der Gibbs-Maschine

Die Gibbs-Maschine kann in der Regel zwei Zielen dienen. Einerseits kann mit ihrer Hilfe aus der Verteilung eines Gibbs-Feldes ein Bild gezogen, andererseits aber auch ein unter der Verteilung maximal wahrscheinliches Bild gesucht werden.

4.3.1 Sampling

Eine Möglichkeit, Computergraphiken ein möglichst realistisches Aussehen zu geben, ist die Darstellung von Oberflächen mittels Texturen (siehe Abschnitt 3.1). Für derartige Zwecke ist ein Mechanismus nötig, der gemäß der Verteilung des zugrundegelegten Gibbs-Feldes eine Realisierung erzeugt.

Die Gibbs-Maschine läßt sich sehr leicht dazu herrichten. Grauwertmenge, Nachbarschaftssystem und Potential sind durch das Zufallsfeld bereits festgelegt. Die Scan-Funktion s kann, wie in allen anderen Fällen auch, als einfacher Raster-Scan gewählt werden, d.h. die Bildpunkte werden zeilenweise von links nach rechts und von oben nach unten durchlaufen:

$$s(t) = \left(1 + (t \bmod N),\, M - \left(\left\lfloor \frac{t}{N} \right\rfloor \bmod M\right)\right) \qquad \forall\, t \in \mathbb{N}_0$$

Wegen $T > 0$ für alle Gibbs-Felder kann das konstante Abkühlschema

$$c(t) = T \qquad \forall\, t \in \mathbb{N}_0$$

verwendet werden (Lemmata 4.6, 4.8). Es hat wesentliche Vorteile bei der Implementation, denn sowohl für MA als auch für GS gilt dann $\Upsilon^{(t)} = \Upsilon^{(t')}$ für alle $t, t' \in \mathbb{N}_0$. Das aber bedeutet, daß die Ersetzungswahrscheinlichkeit für jedes Paar aus Nachbarschaftskonfiguration und Ersetzung höchstens einmal berechnet werden muß, was insbesondere bei homogenen Zufallsfeldern einen immensen Zeitgewinn zur Folge hat.

4.3.2 Optimierung

Bei weitem wichtiger ist der schon mehrfach angesprochene Grenzübergang $T \to 0$. Es gilt nämlich für beliebige Energiefunktionen U

$$\lim_{T \downarrow 0} \frac{1}{Z(T)}\, e^{-U(x)/T} = \begin{cases} \frac{1}{|\hat{\Gamma}|} & x \in \hat{\Gamma} \\ 0 & x \notin \hat{\Gamma} \end{cases} \tag{4.5}$$

mit $\hat{\Gamma} = \{x \in \Gamma^L \,|\, U(x) \leq U(y) \quad \forall\, y \in \Gamma^L\}$. Mit anderen Worten: Die Verteilung eines Gibbs-Feldes konvergiert bei fallender Temperatur gegen eine Gleichverteilung auf der Menge der Bilder mit minimaler Energie. Zum Nachweis von (4.5) sei $\hat{x} \in \hat{\Gamma}$ beliebig, aber fest. Dann gilt

$$\lim_{T \downarrow 0} \sum_{x \in \Gamma^L} \exp\left\{\frac{U(\hat{x}) - U(x)}{T}\right\} = |\hat{\Gamma}|$$

und also

$$\begin{aligned} \lim_{T \downarrow 0} \frac{1}{Z(T)}\, e^{-U(x)/T} &= \lim_{T \downarrow 0} \frac{\exp\{-U(x)/T\}}{\sum_{y \in \Gamma^L} \exp\{-U(y)/T\}} \\ &= \lim_{T \downarrow 0} \frac{\exp\{U(\hat{x}) - U(x)/T\}}{\sum_{y \in \Gamma^L} \exp\{U(\hat{x}) - U(y)/T\}} \\ &= \begin{cases} \frac{1}{|\hat{\Gamma}|} & x \in \hat{\Gamma} \\ 0 & x \notin \hat{\Gamma}. \end{cases} \end{aligned}$$

Beziehung (4.5) bedeutet aber nichts anderes, als daß eine konvergente Gibbs-Maschine zu einem Markoff-Feld mit Grenztemperatur 0 ein Bild mit maximaler Wahrscheinlichkeit, also minimaler Energie, liefert.

Kapitel 3 hat gezeigt, daß Probleme im Rahmen der stochastischen Bildverarbeitung in der Regel durch Schätzer gelöst werden. Die Bestimmung eines solchen wiederum erfolgt durch Argument-Maximierung des zugehörigen Wahrscheinlichkeitsmaßes (Likelihood, *a-posteriori* Verteilung etc.). Um das Problem durch die Gibbs-Maschine zu lösen, ist demnach erstens zu zeigen, daß es sich bei diesem Wahrscheinlichkeitsmaß um eine Gibbs-Verteilung handelt, und zweitens ist deren Potential zu ermitteln.

Mit einem Abkühlschema entsprechend Lemma 4.6 bzw. 4.8 liefert die zugehörige Gibbs-Maschine dann die entsprechende Schätzung. Wegen seiner für praktische Zwecke zu langsamen Konvergenz wird das in den Lemmata angegebene Abkühlschema häufig durch schnellere Varianten ersetzt, für die dann allerdings die Konvergenz der Berechnung nicht mehr gesichert ist. Ein in Verbindung mit dem Raster-Scan gängiges Abkühlschema ist

$$c(t) = \frac{\Delta}{1 + \left\lfloor \frac{t}{N \cdot M} \right\rfloor}$$

mit Δ wie in (4.3).

Abschließend soll ein Beispiel zeigen, wie man ein konkretes Schätzproblem auf die Gibbs-Maschine überträgt.

Beispiel 4.9 (Bildrestauration)
Zur Lösung des Restaurationsproblems aus Abschnitt 3.2 mußte die Energiefunktion (3.6) der a-posteriori *Verteilung minimiert werden. Will man dazu eine Gibbs-Maschine* $\mathcal{G} = (\Gamma, L_F, \tilde{\eta}, \tilde{V}, s, c, \Upsilon)$ *verwenden, so ist im wesentlichen noch die Update-Regel anzugeben. Geman & Geman schlugen dazu den Gibbs-Sampler vor [GG84]. Auf Grund der darin nicht auftretenden Zustandssummen ist der Metropolis-Algorithmus jedoch überschaubarer. Dessen Ersetzungsschritt zum Zeitpunkt* $t \in \mathbb{N}_0$ *hängt nämlich bei aktuellem Bild* (f, e) *und Grauwertvorschlag* $g_{s(t)}$ *lediglich von der Differenz*

$$\begin{aligned}
\Delta_{g_{s(t)}} &= \tilde{U}(f^{g_{s(t)}}, e) - \tilde{U}(f, e) \\
&= U(f^{g_{s(t)}}, e) - U(f, e) \\
&\quad + \frac{\left\| \bar{\mu} - \left(\tau(H(f^{g_{s(t)}})) \ominus \tilde{f}\right)\right\|^2 - \left\|\bar{\mu} - \left(\tau(H(f)) \ominus \tilde{f}\right)\right\|^2}{2\sigma^2} \\
&= \sum_{C \in \mathcal{C}:\, s(t) \in C} V_C(f^{g_{s(t)}}, e) - V_C(f, e) \\
&\quad + \frac{\sum_{(i,j)\in\eta^H_{s(t)}} \left(\mu - \tau\left(\sum_{(k,l)\in\eta^H_{ij}} h_{i-k,j-l} f^{g_{s(t)}}_{kl}\right) \ominus \tilde{f}_{s(t)}\right)^2}{2\sigma^2} \\
&\quad - \frac{\sum_{(i,j)\in\eta^H_{s(t)}} \left(\mu - \tau\left(\sum_{(k,l)\in\eta^H_{ij}} h_{i-k,j-l} f_{kl}\right) \ominus \tilde{f}_{s(t)}\right)^2}{2\sigma^2}
\end{aligned}$$

ab. Am letzten Term wird wegen der geschachtelten Summen noch einmal deutlich, warum η *um* η^{H^2} *zum Nachbarschaftssystem* $\tilde{\eta}$ *des zur* a-posteriori *Verteilung gehörigen Gibbs-Feldes erweitert werden mußte. Außerdem beachte man, daß die Cliquenpotentiale* $\tilde{V}_C$ *nicht explizit berechnet werden müssen. Mit Kenntnis von* $\Delta_{g_{s(t)}}$ *genügt nun die Bestimmung der Akzeptanzwahrscheinlichkeit*

$$\exp\left\{-\frac{\Delta^+_{g_{s(t)}}}{c(t)}\right\}.$$

4.4 Bemerkungen zur Parallelisierung

Ein Arbeitsschritt der GM besteht in der Hauptsache aus der Anwendung der durch die Update-Regel bestimmten Ersetzungsvorschrift. Diese aber hängt nur von der aktuellen Nachbarschaftskonfiguration ab und ist damit *lokal*. Diese Lokalität wiederum ermöglicht die parallele Bearbeitung zumindest von disjunkten geschlossenen Nachbarschaften.

Man kann daher mit Sicherheit die folgende konservative Parallelisierung vornehmen:

Definition 4.10 (Synchrone parallele Gibbs-Maschine)
Es sei X ein Gibbs-Feld über L bezüglich η mit Potential V und Temperatur T. Dann heißt $\mathcal{G}_{par} = (\Gamma, L, \eta, V, S, c, \Upsilon)$ (synchrone) parallele Gibbs-Maschine (PGM) *zu X, falls*

- *$S : \mathbb{N}_0 \to \mathcal{P}(L)$ eine Funktion mit $(k,l) \notin \eta_{ij}$ für alle $(i,j),(k,l) \in S(t)$ und alle $t \in \mathbb{N}_0$ ist, für die $L \subseteq \bigcup_{i=0}^{n-1} S(m+i)$ für ein $n \in \mathbb{N}$ und alle $m \in \mathbb{N}_0$ gilt,*
- *$c : \mathbb{N}_0 \to \mathbb{R}^+$ eine monoton fallende Funktion mit $\lim_{t\to\infty} c(t) = T$ ist und*
- *$\Upsilon_{ij}^{(t)} : \Gamma^{\overline{\eta_{ij}}} \to \{X_{ij}^{(t)} | X_{ij}^{(t)} : (\Omega, \mathcal{A}) \to (\Gamma, \mathcal{P}(\Gamma))\}$ für jedes $t \in \mathbb{N}_0$ und $(i,j) \in S(t)$ eine Funktion ist, die Nachbarschaftskonfigurationen auf Grauwertzufallsvariablen abbildet.*

Dabei heißen S Scan-Funktion, *c* Abkühlschema (cooling schedule) *und* $\Upsilon = \{\Upsilon_{ij}^{(t)} | t \in \mathbb{N}_0, (i,j) \in S(t)\}$ Update-Regel *der PGM.*

Die so definierte parallele GM arbeitet genau wie die sequentielle, mit dem einen Unterschied, daß zu jedem Zeitpunkt durch die Scan-Funktion eine *Menge* von Bildpunkten zur Bearbeitung ausgewählt wird. Die möglichen Auswahlen sind dahingehend beschränkt, daß sich die geschlossenen Nachbarschaften ihrer Elemente nicht überlappen dürfen. Es ist offensichtlich, wie sich MA und GS auf die PGM übertragen und auch, daß die Konvergenzresultate erhalten bleiben[3].

Gemäß der von Greening [Grg89] angegebenen Taxonomie der Algorithmen für paralleles Simulated Annealing ist die PGM zu den synchronen datenparallelen Verfahren (*spatial decomposition with independent processors*) zu zählen.

In der Synchronisation der Ersetzungen durch die Scan-Funktion, die gewissermaßen den Takt des Algorithmus' vorgibt, besteht aber auch die inhärente

[3] Es reicht aus, die Scan-Funktion S der PGM in eine Scan-Funktion s für eine GM zu zerlegen, also z.B.:

$$\overbrace{s(1)\,s(2)\,s(3)\,s(4)}^{S(1)}\overbrace{s(5)\,s(6)\,s(7)\,s(8)}^{S(2)}\ldots$$

Einschränkung der PGM. Wünschenswerter wäre es, die Ersetzungen an den Bildpunkten *asynchron* durchzuführen, weil dann auf die rechenzeitintensive Abstimmung der Prozessoren verzichtet werden könnte.

Da nicht anzunehmen ist, daß genügend Recheneinheiten zur Verfügung stehen, um jedem Bildpunkt einen eigenen Prozessor zuzuweisen, kann man davon ausgehen, daß kein Grauwert zweimal aktualisiert wird, während die Berechnung für einen anderen noch läuft. Prinzipiell läßt sich daher auch für einen asynchron arbeitenden Algorithmus eine Scan-Funktion (sozusagen im nachhinein) angeben, wenn man von der Forderung nach disjunkten Nachbarschaften absieht. Eine asynchron parallele GM entspricht dann einer PGM mit uneingeschränkter Scan-Funktion, d.h. $S : \mathbb{N}_0 \to \mathcal{P}(L)$ ist beliebig.

Es ist aber nicht bekannt [AK89, Grg89], ob oder unter welchen Bedingungen die gelegentlich konfliktbelastete Aktualisierung von Bildpunkten ohne Auswirkungen auf die Konvergenz bleibt. Wird nämlich der Versuch unternommen, die Grauwerte zweier benachbarter Bildpunkte gleichzeitig zu aktualisieren, kann es zu unerwünschten Effekten kommen: Möglicherweise haben beide Änderungen für sich genommen jeweils eine Verringerung der Energie zur Folge, bewirken zusammen aber eine Erhöhung. Lediglich für die gleichzeitige Aktualisierung aller Pixel (volle synchrone Parallelität), d.h. $S(t) = L$ für alle $t \in \mathbb{N}_0$, weiß man, daß für $T \to 0$ nicht mehr U, sondern die Pseudo-Energie

$$\bar{U}(x) = \sum_{(i,j)\in L} \log \left(\sum_{y_{ij}\in\Gamma} \exp \left\{ \sum_{C\in\mathcal{C}:(i,j)\in C} V_C(x^{y_{ij}}) \right\} \right)$$

minimiert wird [Aze90, Ruf93].

4.5 Parameterschätzung

In Abschnitt 3.1.3 wurde die MPL-Schätzung der Parameter einer Gibbs-Verteilung mit parametrisiertem Potential vorgestellt. Obwohl Lakshmanan und Derin ihre in [LD89] veröffentlichte Methode zur Bestimmung einer MPL-Schätzung so nennen, handelt es sich im strengen Sinne nicht um Simulated Annealing, weil ihr Parameterraum überabzählbar ist.

Da aber bei der Implementation der Verfahren durch den begrenzten Speicherplatz im Rechner ohnehin nur ein endlicher Parameterraum abgesucht wird, kann ein der Gibbs-Maschine analoges Maschinenmodell angegeben werden, das eine nur geringfügig abgewandelte Form des Verfahrens formalisiert.

Definition 4.11 (Duale Gibbs-Maschine)
Es sei X ein Gibbs-Feld über L bezüglich η mit endlich parametrisiertem Potential V. Dann heißt $\mathcal{G}^* = (\Gamma^*, L^*, V, s^*, c, \Upsilon^*)$ duale Gibbs-Maschine (DGM) *zu X, falls*

- *$s^* : \mathbb{N}_0 \to L^*$ eine Funktion mit $L^* \subseteq \{s^*(m), \ldots, s^*(m+n-1)\}$ für ein $n \in \mathbb{N}$ und alle $m \in \mathbb{N}_0$ ist,*
- *$c : \mathbb{N}_0 \to \mathbb{R}^+$ eine monoton fallende Funktion mit $\lim_{t\to\infty} c(t) = 0$ ist und*
- *$\Upsilon^{*(t)} : L^* \to \{\Theta^{(t)}_{s^*(t)} | \Theta^{(t)}_{s^*(t)} : (\Omega, \mathcal{A}) \to (\Gamma^*_{s^*(t)}, \mathcal{P}(\Gamma^*_{s^*(t)}))\}$ für jedes $t \in \mathbb{N}_0$ eine Funktion ist, die Parameterindizes auf Parameterzufallsvariablen abbildet.*

Dabei heißen s^* Scan-Funktion, *c* Abkühlschema (cooling schedule) *und $\Upsilon^* = \{\Upsilon^{*(t)}\}_{t\in\mathbb{N}_0}$* Update-Regel *der DGM.*

Analog zu (4.1) induziert auch die duale Gibbs-Maschine eine Markoff-Kette zur Beschreibung der möglichen Berechnungen $\{\theta^{(t)}\}_{t\in\mathbb{N}_0}$. Mit der logarithmierten Pseudo-Likelihood

$$U^*(\theta) := \log \prod_{(i,j)\in L} P_\theta(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}})$$

kann die Update-Regel MA für die DGM in der gewohnten Form angegeben werden. Dazu sei $\theta^{(t)} = \theta$ die aktuelle Parameterschätzung zum Zeitpunkt $t \in \mathbb{N}_0$ und $\theta' = \theta^{\theta'_{s^*(t)}}$ der höchstens an der Stelle $s^*(t)$ gegenüber θ geänderte Vorschlag für den Folgezustand. Dann ist bei MA die Wahrscheinlichkeit dafür, daß $\theta_{s^*(t)}$ durch $\theta'_{s^*(t)}$ ersetzt wird

$$P\left(\Upsilon^{*(t)}(s^*(t)) = \theta'_{s^*(t)}\right)$$
$$= \begin{cases} \frac{1}{|\Gamma^*_{s^*(t)}|} \exp\left\{-\frac{(U^*(\theta)-U^*(\theta'))^+}{c(t)}\right\} & \theta'_{s^*(t)} \neq \theta_{s^*(t)} \\ 1 - \sum\limits_{\substack{g^*_{s^*(t)}\in\Gamma^*_{s^*(t)} \\ g^*_{s^*(t)}\neq\theta_{s^*(t)}}} \frac{1}{|\Gamma^*_{s^*(t)}|} \exp\left\{-\frac{(U^*(\theta)-U^*(\theta^{g^*_{s^*(t)}}))^+}{c(t)}\right\} & \theta'_{s^*(t)} = \theta_{s^*(t)}. \end{cases}$$

In einer konkreten Implementation entstehen diese Wahrscheinlichkeiten, indem zunächst gleichverteilt ein möglicher Nachfolgewert $\theta'_{s^*(t)}$ für die Parameterkomponente mit Index $s^*(t)$ generiert wird. Anschließend wird der Quotient

$$q = \left(\frac{\prod_{(i,j)\in L} P_{\theta'}(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}})}{\prod_{(i,j)\in L} P_\theta(X_{ij} = x_{ij} | X_{\eta_{ij}} = x_{\eta_{ij}})} \right)^{\frac{1}{c(t)}}$$

bestimmt und $\theta'_{s^*(t)}$ akzeptiert, falls $q \geq 1$. Ansonsten wird der neue Wert nur mit Wahrscheinlichkeit q übernommen. Man vergewissere sich, daß dies der Definition von MA für die DGM entspricht.

Für inhomogene Markoff-Felder ist die Berechnung von q lokal auf L^*, wodurch eine Parallelisierung ähnlich der PGM möglich wird. Doch bei den wichtigeren homogenen MRFs ist die Berechnung von q weder auf L noch auf L^* lokal, da die Zustandssummen der einzelnen Faktoren über alle zulässigen Grauwerte am Punkt (i,j) summieren und deshalb bezüglich der Parameter nicht separabel sind. In diesem Fall wäre daher lediglich eine parallele DGM möglich, welche die Berechnung der einzelnen Faktoren von q verteilt. In der erwähnten SA-Taxonomie Greenings [Grg89] gehörte eine derartige Implementation des SA zu den Algorithmen mit Funktionsdekomposition (*functional decomposition algorithms*).

Bei sehr großen, abzählbaren oder gar überabzählbaren Parameterräumen empfiehlt sich zur Vermeidung zu großer Sprünge eine andere Generierung der möglichen Nachfolgezustände. Unter Berücksichtigung von Randbedingungen kann ein Änderungsvorschlag für $\theta_{s^*(t)}$ mit hinreichend kleinem ϵ gleichverteilt auf dem Intervall $[\theta_{s^*(t)} - \epsilon, \theta_{s^*(t)} + \epsilon]$ gezogen werden [LD89].

Kapitel 5

Schlußbemerkungen

> "Und der etwas einfältige Stil, in dem sie geschrieben sind, ist zum Teil dadurch zu erklären, daß die Herausgeber, die den Erscheinungstermin des Buches einhalten wollten, alle Informationen von der Rückseite eines Packets Frühstücksflocken abschrieben (...)."
>
> D. ADAMS: DAS RESTAURANT AM ENDE DES UNIVERSUMS

In dieser Ausarbeitung wurden die aus der Literatur bekannten Markoff-Felder als stochastisches Modell für Digitalbilder vorgestellt. Nach Einführung über ihre lokalen Charakteristiken (die Markoff-Eigenschaft) wurde gezeigt, daß es zu jedem Markoff-Feld eine korrespondierende globale Beschreibung durch eine Gibbs-Verteilung gibt. In beiden Fällen wurde besonderer Wert auf die üblicherweise fehlende Motivierung der Ansätze gelegt. Des weiteren wurde der Begriff des parametrisierten Potentials konkretisiert und eine anschaulich interpretierbare Parametrisierung für eine neue Teilklasse von praktischer Relevanz angegeben.

Verbreitung und Bedeutung der Markoff-Felder in der Bildverarbeitung wurden durch einige Anwendungsbeispiele angedeutet. Anhand konkreter Aufgabenstellungen wurden die unmittelbar plausiblen Vorteile der Modellierung durch lokale Wechselwirkungen untermauert [ChJ93, Vorwort]:

- *Systematischer Entwurf von Algorithmen:* Das wohldefinierte Schema gestattet gegenüber der häufig heuristischen modellfreien Verfahrensweise ein fundierteres Vorgehen. Man kann sogar hoffen [GMcC85], eine noch konsequentere Vereinheitlichung zahlreicher Bildverarbeitungsprobleme durch Rückführung auf das Maximum *a-posteriori* Schätzproblem zu erreichen (in Abschnitt 3.5 begonnen).

- *Bewertung:* Durch die Formulierung als Optimierungsaufgabe ist in den Algorithmen meist erkennbar (und in vielen Fällen sogar quantifizierbar), an welchen Stellen der Algorithmen lediglich Annäherungen an die optimale Lösung bestimmt werden.

- *Flexibilität:* Die in der MAP-Schätzung auftretenden *a-priori* Verteilungen gestatten es, vorhandenes Vorwissen oder Randbedingungen in die Modellierung zu integrieren.

- *Parallelität:* Die inhärente Lokalität des Markoff-Modells bedingt fast automatisch eine gute und leicht zu nutzende Parallelisierbarkeit der Berechnungen.

Der Entwurf eines allgemeinen Berechnungsmodells in Form der Gibbs-Maschine führte die im Anschluß an die Anwendungsbeispiele begonnene Vereinheitlichung der Problemlösungen fort. Dieses Maschinenmodell soll dazu beitragen, algorithmische Untersuchungen, insbesondere Parallelisierungs- und Konvergenzfragen, von den eigentlichen Problemstellungen losgelöst durchführen zu können. Weiterhin konnte gezeigt werden, daß sich das Modell auch für die Schätzung von Parametern modifizieren läßt.

Es wurde bereits eingangs die Auffassung vertreten, daß Bildverarbeitung nicht bloß die Behandlung visueller Information betrifft. Die vorgestellten Konzepte lassen sich ohne weiteres auf andere Gebiete übertragen, denn auch räumliche Informationen über Infektion von Nutzpflanzenkulturen, Abwasserverschmutzung, Wolkenformationen, Gasdichten, Molekularzustände und vieles andere mehr können als "Bilder" angesehen werden.

Diese breite Anwendbarkeit der Markoff-Felder rührt unter anderem daher, daß die Ausdehnung des unterlegten Gitters auf beliebige Dimensionen ohne jede Einschränkung möglich ist, da Nachbarschaften lediglich Teilmengen der Gitterpunkte, also dimensionslos sind. Ebenso spielt es für die Modellebene keine Rolle, ob die Komponenten der Zufallsfelder kontinuierlich und/oder vektorwertig sind. Ein Beispiel für Gittermodelle mit kontinuierlichen Komponenten sind die Gauß-Markoff-Felder (GMRF, [ChJ93] enthält mehrere Kapitel darüber). Der dann überabzählbare Bildraum macht allerdings Modifikationen des Berechnungsmodells notwendig [ChJ93, Kapitel 18].

Neben den zahlreichen Spielarten und Erweiterungen der Markoff-Felder existieren noch ganz andere Klassen von stochastischen Bildmodellen. Einen Überblick gibt Jain in [Jai81].

Einer der sich aus dem Markoff-Modell ergebenden Nachteile ist der – wegen der im allgemeinen langsamen Konvergenz der stochastischen Algorithmen – hohe Berechnungsaufwand. Der Grund liegt darin, daß sich Abhängigkeiten über große Entfernungen auswirken können, während der Berechnungen aber nur lokaler Einfluß genommen werden kann. Von den zahlreichen verbleibenden Aufgaben zählen daher zu den interessantesten

- die Beantwortung der Frage nach der Konvergenz asynchron paralleler Gibbs-Maschinen und
- die weitergehende Untersuchung der *multiresolution*-Modelle (Überblick in [ChJ93, Kapitel 6]).

Nutzbringend wären aber auch neue Erkenntnisse auf dem Gebiet

- der Parameterschätzung, denn eine günstige *a-priori* Verteilung trägt wesentlich zur Qualität der Schätzresultate bei (siehe beispielsweise Abschnitt 3.4),
- der Kombination von Bildern unterschiedlicher Sorten von Information (wie etwa bei Lipinski [Lip94], vgl. Abschnitt 3.4.4) und
- der Integration von (eventuell hierarchischen) abstrakten Bildinhalten (wie etwa der Kantenelemente aus Abschnitt 3.2).

Trotz der verschiedenen Modellierungsansätze und Erweiterungsmöglichkeiten ist nicht anzunehmen, daß eine objektive (quantifizierbare) Güteaussage über die Qualität der Ergebnisse, insbesondere im Vergleich mit den Verfahren der deterministischen Bildverarbeitung, möglich werden wird. Denn obwohl die Lösung der Probleme in der Regel aus einem (stochastisch) optimalen Schätzer besteht, gilt diese Optimalität doch nur innerhalb des jeweiligen Modells. Und über dessen Adäquatheit läßt sich, seinem anschaulichen Reiz zum Trotz, trefflich streiten.

Anhang A

Implementation

Die in Abschnitt 3.1.2 gezeigten Beispiele für Texturen wurden durch das aus diesem Anlaß entstandene Programm `gm` generiert, das eine praktische Umsetzung der in Kapitel 4 eingeführten Gibbs-Maschine darstellt. Es ist in C geschrieben und verwendet die SIMPLEX-Bibliothek zur Einbindung in die UNIX-Umgebung. Abbildung A.1 zeigt die Programmoberfläche.

Das Programm ist ausgelegt für homogene Kontrast-Modelle (Abschnitt 2.5) über den Nachbarschaftssystemen η^0 bis η^4 der hierarchischen Folge aus Beispiel 2.5 (v). Die Grauwertparameter werden für alle $g \in \Gamma$ unveränderlich $\alpha_g = 0$ gesetzt, die Bindungsstärken $\beta_{ij,kl}$ können im unteren linken Bereich durch Eingabe über die Tastatur modifiziert werden. Die Anfangstemperatur $c(0)$ kann ebenfalls vorgegeben werden. Die Anzahl G der Grauwerte, die Größe $N \cdot M$ des Gitters L und die Kontrastfunktion $\delta : \Gamma \times \Gamma \to [0, 1]$ sind im Programmtext wie folgt festgelegt (und nur dort zu ändern):

```
#define G   32
#define N  128
#define M  128

double delta(int g,h)
{
  return fabs(g-h)/(G-1);
}
```

Aus optisch-ästhetischen Gründen wird der nächste zu aktualisierende Punkt nicht durch eine Scan-Funktion, sondern durch eine auf dem Gitter gleichverteilte Pseudo-Zufallsfunktion bestimmt. Die einem einzelnen Raster-Scan entsprechende Anzahl $N \cdot M$ von Update-Schritten stellt eine Iteration der Berechnung dar. Als Update-Regel wurde MA gewählt, weswegen eine Iteration der Gibbs-Maschine durch die folgende Prozedur realisiert wird:

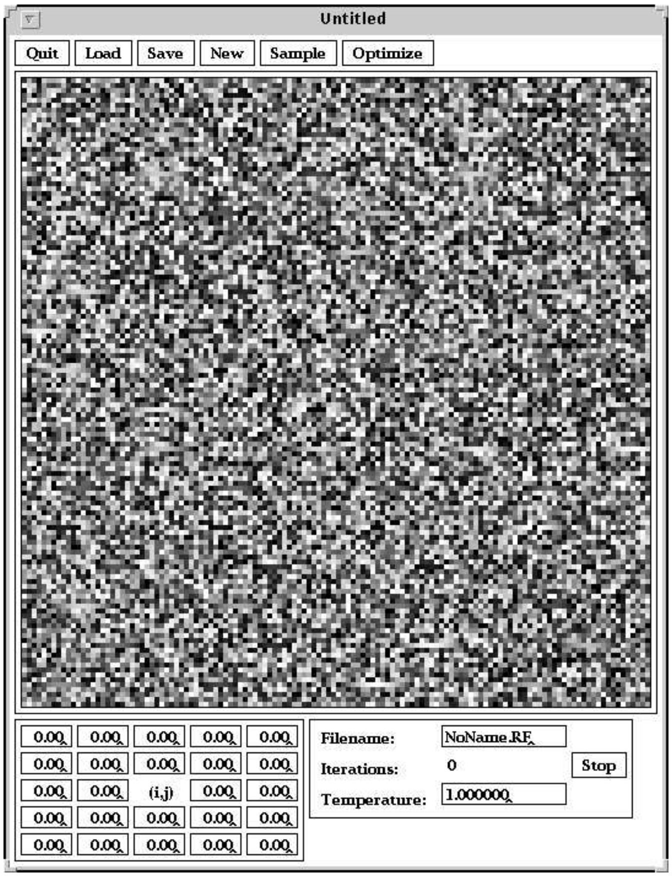


Abbildung A.1: Benutzerschnittstelle des Programms `gm`

```
void sweep(float T)
{
  int     i,j,l,g;
  double  diff;

  for(l=0;l<N*M;l++){
    i = rand()%N;  j = rand()%M;     /* random pixel      */
    g = rand()%G;                    /* random grey value */
    diff = energy_difference(i,j,g); /* energy change     */
    if ((diff<=0) || (exp(-diff/T) < random(0,1)))
      x[i][j] = g;                   /* accept */
  }
}
```

Dabei speichert die globale Variable `x[N][M]` das jeweils aktuelle Bild. Die durch die Optionen der über dem Bildbereich befindlichen Menüzeile ausgelösten Aktionen sind:

- `Quit`: Beenden des Programms.
- `Load`: Laden einer Parametrisierung aus der im Feld `Filename`: spezifizierten Datei.
- `Save`: Speichern der aktuellen Parametrisierung in der im Feld `Filename`: spezifizierten Datei.
- `New`: Rücksetzen aller Bindungsstärken $\beta_{ij,kl}$ auf den Wert 0 und der Temperatur auf den Wert $T = 1$. Erzeugung eines zufälligen Anfangsbildes.
- `Sample`: Generierung einer Textur mit der aktuellen Parametrisierung bei der angegebenen Temperatur. Als Abkühlschema wird die konstante Funktion $c(t) = T$ für alle $t \in \mathbb{N}_0$ verwendet.
- `Optimize`: Generierung einer Näherung für ein unter der aktuellen Parametrisierung maximal wahrscheinliches Bild. Es wird mit einem als Konstante `GAMMA` vereinbarten Parameter $\gamma \in (0,1)$ nach jeder Iteration geometrisch abgekühlt, d.h.

$$c(t+1) = \begin{cases} \gamma \cdot c(t) & \text{falls } t+1 \equiv 0 \bmod (N \cdot M) \\ c(t) & \text{sonst.} \end{cases}$$

Sowohl beim Sampling als auch beim Optimieren können die aktuelle Parametrisierung und die angezeigte Temperatur während der Berechnung beeinflußt werden.

Anhang B

Symbolverzeichnis

$\mathbb{N}$	Menge der natürlichen Zahlen $\{1, 2, 3, \ldots\}$	
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$	
$\mathbb{Z}$	Menge der ganzen Zahlen	
$\mathbb{R}$	Menge der reellen Zahlen	
$\emptyset$	Leere Menge	
$\mathcal{P}(A)$	Potenzmenge von A	
$\prod_{i=1}^{n} A_i$	Kartesisches Produkt $A_1 \times \ldots \times A_n$	33
$\{a^{(t)}\}_{t \in \mathbb{N}_0}$	Folge $a^{(0)}, a^{(1)}, a^{(2)}, \ldots$	
a^+	Positivteil der reellen Zahl a	91
$\delta_{a,b}$	Kronecker-Symbol	
X, Y	Zufallsfelder $(\Omega, \mathcal{A}) \to (\Gamma^L, \mathcal{P}(\Gamma^L))$	8
$P^X(x) = P(X = x)$	Induziertes Wahrscheinlichkeitsmaß	
$X \sim \mathcal{POISSON}(\lambda)$	Mit Parameter λ Poisson-verteilte ZV'e X	
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung	
$E(X)$	Erwartungswert der ZV'e X	
$P(X = x \mid Y = y)$	Bedingte Wahrscheinlichkeit	
$X_\Lambda, x_\Lambda, \theta_\Lambda$	Teilvektoren	15
$x^{\{y_{kl} \mid (k,l) \in \Lambda\}} = x^{y_\Lambda}$	Modifiziertes Bild	26
$I = \{1, \ldots, N\}$	Menge der Zeilenindizes	1
$J = \{1, \ldots, M\}$	Menge der Spaltenindizes	1
$L = I \times J$	Gitter	1
$L_{d_i d_j}$	Teilgitter (Definitionsbereich von $\tau_{d_i d_j}$)	16
$\tau_{d_i d_j}$	Gittertranslation	16
η_{ij}	Nachbarschaft von $(i, j) \in L$	12
$\overline{\eta_{ij}}$	Geschlossene Nachbarschaft	12

η	Nachbarschaftssystem	12
$\eta^0, \eta^1, \ldots$	Spezielle Nachbarschaftssysteme	12
C	Clique	20
$\mathcal{C} = \mathcal{C}(\eta)$	Menge der Cliquen über η	20
G	Anzahl der Grauwerte	1
$\Gamma = \{0, \ldots, G-1\}$	Menge der Grauwerte	1
Γ^L	Bildraum	1
L^*	Duales Gitter (Parameterindizes)	33
Γ^*	Parameterraum	33
θ	Parametrisierung	33
$V(\theta)$	Parametrisiertes Potential	33
$U(\theta)$	Parametrisierte Energie	33
P_θ	Parametrisierte Verteilung	33
Z^θ, Z^θ_{ij}	Zustandssummen einer parametrisierten Gibbs-Verteilung	33
$\delta : \Gamma^2 \rightarrow [0,1]$	Kontrastfunktion	34
α_g	Einercliquengewicht in Auto-Modellen	32
$\beta_{ij,kl}$	Zweiercliquengewicht (Bindunsgstärke)	32
$\beta_1, \ldots, \beta_4$	Bindungstypen in η^2	36
V_C	Potential der Clique C	21
V	Menge der Cliquenpotentiale	21
$U = U_V$	Energie des Potentials V	21
T	Temperatur	23
$Z_{ij}(T)$, Z_{ij}	Lokale Zustandssumme	26
$Z(T)$, Z	Zustandssumme	23
$\mathcal{G}$, $\mathcal{G}_{par}$, $\mathcal{G}^*$	Gibbs-Maschine (GM, PGM bzw. DGM)	87,98,100
s, S, s^*	Scan-Funktion der GM, PGM bzw. DGM	87,98,100
c	Abkühlschema	87,98,100
Υ, Υ^*	Update-Regeln	87,98,100

Literaturverzeichnis

[AKM93] T. Aach, A. Kaup, R. Mester: *Statistical model-based change detection in moving video*, Signal Processing, vol. 31, 1993, pp. 165–180

[AK89] E. H. L. Aarts, J. Korst: *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons Ltd., New York, 1989

[AHK65] K. Abend, T. J. Harley, L. N. Kanal: *Classification of binary random patterns*, IEEE Transactions on Information Theory, vol. IT-11, no. 4, 1965, pp. 538–544

[Acu92] C. O. Acuna: *Texture modelling using Gibbs distributions*, CVGIP: Graphical Models and Image Processing, vol. 54, no. 3, 1992, pp. 210–222

[Aig93] M. Aigner: *Diskrete Mathematik*, Vieweg, Berlin, 1993

[Ave70] M. B. Averintsev: *One method of describing random fields with a discrete parameter*, Problems of Information Transmission, Januar 1973, pp. 169-175; Engl. Übersetzung des Originals aus: Problemy Peredači Informacii, vol. 6, no. 2, 1970, pp. 100–109

[Aze90] R. Azencott: *Synchronous Boltzmann machines and Gibbs fields: Learning algorithms*; in: *Neurocomputing*, F. Fogelman Soulié, J. Hérault (eds.), NATO ASI Series, vol. F68, Springer, Berlin, 1990

[Bar55] M. S. Bartlett: *An Introduction to Stochastic Processes*, Cambridge: University Press, 1955

[Bar67] M. S. Bartlett: *Inference and stochastic processes*, Journal of the Royal Statistical Society, vol. A-130, 1967, pp. 457–477

[Bar68] M. S. Bartlett: *A further note on nearest neighbour models*, Journal of the Royal Statistical Society, vol. A-131, pp. 579–580

[Bat86] R. H. T. Bates, M. J. McDonnel: *Image Restoration and Reconstruction*, Oxford: University Press, 1986

[Bes74] J. Besag: *Spatial interaction and the statistical analysis of lattice systems* (mit Diskussion), Journal of the Royal Statistical Society, vol. B-36, no. 2, 1974, pp. 192–236

[Bes86] J. Besag: *On the statistical analysis of dirty pictures*, Journal of the Royal Statistical Society, vol. B-48, no. 3, 1986, pp. 259–279

[Bro66] P. Brodatz: *Textures – A Photographic Album for Artists & Designers*, Dover Publications, New York, 1966

[Cha88] B. Chalmond: *Image restoration using an estimated Markov model*, Signal Processing, vol. 15, 1988, pp. 115–129

[Cha89] B. Chalmond: *An iterative Gibbsian technique for reconstruction of m-ary images*, Pattern Recognition, vol. 22, no. 6, 1989, pp. 747–761

[ChJ93] R. Chellappa, A. K. Jain (eds.): *Markov Random Fields: Theory and Applications*, Academic Press, San Diego, 1993

[CrJ83] G. R. Cross, A. K. Jain: *Markov random field texture models*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-5, no. 1, 1983, pp. 25–39

[DLR77] A. P. Dempster, N. M. Laird, D. B. Rubin: *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society, vol. B-39, 1977, pp. 1–38

[DECG84] H. Derin, H. Elliott, R. Cristi, D. Geman: *Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, 1984, pp. 707–720

[Der85] H. Derin: *The use of Gibbs distributions in image processing*; in: *Communications and Networks: A Survey of Recent Advances*, I. Blake and V. Poor (eds.), Springer Verlag, New York, 1985

[DC86] H. Derin, W. S. Cole: *Segmentation of textured images using Gibbs random fields*, Computer Vision Graphics and Image Processing, vol. 35, 1986, pp. 72–98

[DE87] H. Derin, H. Elliott: *Modeling and segmentation of noisy and textured images using gibbs random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-9, no.1, 1987, pp. 39–55

[Der91] H. Derin: *Commments on "Restoration of noisy images modeled by Markov random fields with Gibbs distribution"* ([SAG89]; mit Entgegnung), IEEE Transactions on Circuits and Systems, vol. 38, no. 5, 1991, pp. 566–568

[Dob68] R. L. Dobrushin: *Description of a random field by means of conditional probabilities and the conditions governing its regularity*, Theory of probability and its applications, vol. 13, 1968, pp. 197–224

[GG84] S. Geman, D. Geman: *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, 1984, pp. 721–741

[GMcC85] S. Geman, D. E. McClure: *Bayesian image analysis: An application to single photon emission tomography*, Proceedings of the American Statistical Association, Statistical Computing Section, 1985, pp. 12–18

[GMcC87] S. Geman, D. E. McClure: *Statistical methods for tomographic image reconstruction*, Proceedings of the ISI Tokio Session, Bulletin of the International Statistical Institute, vol. 52, 1987, pp. 5–21

[GDO85] L. van Gool, P. Dewaele, A. Oosterlinck: *Texture analysis anno 1983*, Computer Vision, Graphics, and Image Processing, vol. 29, 1985, pp. 336–357

[Gre90] P. J. Green: *Bayesian reconstruction from emission tomography data using a modified EM algorithm*, IEEE Transactions on Medical Imaging, vol. 9, no. 1, 1990, pp. 84–93

[Grg89] D. R. Greening: *A taxonomy of parallel simulated annealing techniques*, Research Report 66639, IBM T. J. Watson Research Center, Yorktown Heights, New York, 1989

[GNS87] W. Greiner,L. Neise, H. Stöcker: *Theoretische Physik 9: Thermodynamik und Statistische Mechanik*, Verlag Harri Deutsch, Thun, 1987

[Grif76] D. Griffeath: *Introduction to Random Fields*; in: *Denumerable Markov Chain*, J. Kemeny, L. Snell, A. Knapp (eds.), Springer, New York, 1976, pp. 425–458

[Grim72] G. R. Grimmett: *A theorem about random fields*, Bulletin of the London Mathematical Society, vol. 5, 1972, pp. 81–84

[Hab87] P. Haberäcker: *Digitale Bildverarbeitung*, 2. Aufl., Carl Hanser Verlag, München, 1987

[HH64] J. M. Hammersley, D. C. Handscomb: *Monte Carlo Methods*, Methuen, London, 1964

[Har79] R. M. Haralick: *Statistical and Structural Approaches to Texture*, Proceedings of the IEEE, vol. 67, no. 5, 1979, pp. 786–804

[HS80] M. Hassner, J. Sklansky: *The use of Markov random fields as models of texture*, Computer Graphics and Image Processing, vol. 12, 1980, pp. 357–370

[Hun77] B. R. Hunt: *Bayesian methods in nonlinear digital image restoration*, IEEE Transactions on Computers, vol. C-26, no. 3, 1977, pp. 219–229

[Isi25] E. Ising: *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift Physik, Vol. 31, 1925, pp. 253–258

[Jäh91] B. Jähne: *Digitale Bildverarbeitung*, 2. Aufl., Springer, Berlin, 1991

[JA74] A. K. Jain, E. Angel: *Image restoration, modelling, and reduction of dimensionality*, IEEE Transactions on Computers, vol. C-23, no. 5, 1974, pp. 470–476

[Jai81] A. K. Jain: *Advances in mathematical models for image processing*, Proc. IEEE, vol. 69, 1981, pp. 502–528

[Jul75] B. Julesz: *Experiments in the Perception of Texture*, Scientific American, vol. 232, no. 4, 1975, pp. 34–43

[Kan80] L. N. Kanal: *Markov mesh models*, Computer Graphics and Image Processing, vol. 12, 1980, pp. 371–375

[Kre90] U. Krengel: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 2., verbesserte Auflage, Vieweg, Braunschweig, 1990

[LD89] S. Lakshmanan, H. Derin: *Simultaneous parameter estimation and segmentation of Gibbs random fields using Simulated Annealing*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 8, 1989, pp. 799–813

[Lan90] K. Lange: *Convergence of EM image reconstruction algorithms with Gibbs smoothing*, IEEE Transactions on Medical Imaging, vol. MI-9, no. 4, 1990, pp. 439–446; Korrekturen dazu: vol. MI-10, no. 2, 1991, p. 228

[Lip94] B. Lipinski: *Iterative Rekonstruktion von PET-Bildern unter Nutzung anatomischer MR-Information*, vorgesehener Dissertations-Titel, Aachen, 1994

[Lis86] J. Lissner: *Radiologie I*, 3., neu bearbeitete Auflage, Ferdinand Enke Verlag, Stuttgart, 1986

[MP90] R. Mathar, D. Pfeifer: *Stochastik für Informatiker*, Teubner, Stuttgart, 1990

[MRT53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller: *Equations of state calculations by fast computing machines*, The Journal of Chemical Physics, vol. 21, no. 6, 1953, pp. 1087–1092

[OHW87] W. Oberschelp, G. Hinsen, F. Wagner: *Kombinatorik*, RWTH Aachen, Schriften zur Informatik und Angewandten Mathematik, Nr. 129, 1987

[OD91] W. Oberschelp, M. Dohmen: *Mathematische Methoden für Bildverarbeitung und Computergraphik. Spezielle algorithmische Probleme der Computergraphik*, RWTH Aachen, Schriften zur Informatik und Angewandten Mathematik, Nr. 149, 1991

[RK82] A. Rosenfeld, A. C. Kak: *Digital Picture Processing*, vol. I, 2nd ed., Academic Press, New York, 1982

[Rue68] D. Ruelle: *Statistical Mechanics*, W. A. Benjamin, New York, 1968

[Ruf93] B. Ruf: *Sequentielle und parallele Lernverfahren für Boltzmann-Maschinen*, RWTH Aachen, Fachgruppe Informatik, Diplomarbeit 33/93, 1993

[SV82] L. A. Shepp, Y. Vardi: *Maximum likelihood reconstruction for emission tomography*, IEEE Transactions on Medical Imaging, vol. MI-1, no. 2, 1982, pp. 113–122

[SAG89] M. Shridar, M. Ahmadi, M. El-Gabali: *Restoration of noisy images modeled by Markov random fields with Gibbs distribution*, IEEE Transactions on Circuits and Systems, vol. 36, no. 6, 1989, pp. 884–890

[Sol71] K. D. Solf: *Fotografie*, Fischer Taschenbuch Verlag, Frankfurt am Main, 1971

[Spi71] F. Spitzer: *Markov random fields and Gibbs ensembles*, American Mathematical Monthly, vol. 78, 1971, pp. 340–368

[SK76] J. A. Stuller, B. Kurz: *Two-dimensional Markov representations of sampled images*, IEEE Transactions on Communications, vol. COM-24, no. 10, 1976, pp. 1148–1152

[Tan93] M. A. Tanner: *Tools for Statistical Inference*, Springer Series in Statistics, 2nd ed., New York, 1993

[TB86] P. Thurn, E. Bücheler: *Einführung in die radiologische Diagnostik*, Georg Thieme Verlag, Stuttgart, New York, 1986

[Wei93] F. Weiler: *Merkmalsextraktion aus Filmröntgenbildern zur Abschätzung der Dignität fokaler Knochenläsionen*, RWTH Aachen, Fachgruppe Informatik, Diplomarbeit 31/93, 1993

[Whi63] P. Whittle: *Stochastic processes in several dimensions*, Bulletin of the International Statistical Society, vol. 40, 1963, pp. 974–994

[Win90] K. zum Winkel: *Nuklearmedizin*, 2. Auflage, Springer Verlag, Berlin, 1990

[Woo72] J. W. Woods: *Two-dimensional discrete Markovian fields*, IEEE Transactions on Information Theory, vol. IT-18, no. 2, 1972, pp. 232–240