

# Einführung in die Bayessche Bildanalyse

## Seminar: Bayessche Ansätze in der Bildanalyse

Benedikt Kramps

Fakultät für Mathematik und Wirtschaftswissenschaften  
Universität Ulm

8.Mai 2006



- 1 Einführung
  - Motivation
  - Beispielbilder
- 2 Bilder
  - Computergrafiken
  - Bildarten
- 3 Bayessche Bildanalyse
  - Grundlagen
  - Bayes'sches Paradigma
  - Bayes-Schätzer
- 4 MCMC
  - Grundlagen
  - Eigenschaften
  - Beispiel
  - Algorithmen
- 5 Anhang
  - Literatur

# Bildanalyse

- 'inverse optics'
- inverse Probleme sind unterbestimmt
- verschiedene Interpretationen zulässig
- 'Kunst' die richtige auszuwählen
- Bayes'sche Bildanalyse: stochastischer Zugang
- 'direktes Problem' wird interpretiert als Simulation von Mustern (MCMC)

aus *G. Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer, 2nd ed., 2003.*

# Bildanalyse

- interpretiert den Inhalt von Bildern
- beschreibt elektronische Bildverarbeitung und automatische Mustererkennung
- findet Einsatz in Medizin, Biologie, Geowissenschaften, im räumlichen Marketing und in anderen Bereichen

# Ergebnisse

- numerische oder geometrische Merkmale (Flächeninhalt eines Bereiches)
- topologische oder stereologische Merkmale (Anordnung von Objekten in einer Szene)
- Texturmerkmale (Struktur einer Oberfläche)

# Anwendungsbeispiele

- Bildrekonstruktion aus verrauschten oder anders gestörten Daten
- Segmentierung
- Kantenbestimmung (boundary detection)
- Tomographie (3D aus 2D)
- Transformationsfeld (deformation field) zwischen zwei Bildern
- Bewegungsanalyse (motion analysis)

# Kantenextraktion



Abbildung: links: Original; rechts: Kantenextraktion mit einem Modell von Geman und Geman

# Segmentierung



Abbildung: links: Luftbild von Dessau; rechts: Segmentierung mit einem Region-Growing-Verfahren



# Deformationsfeld

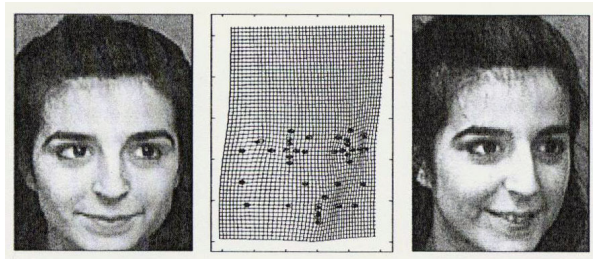
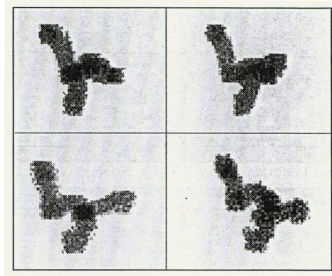


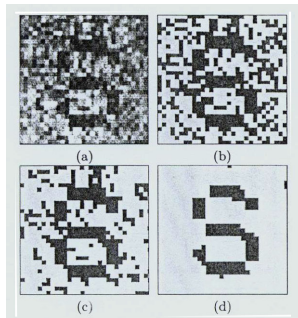
Abbildung: Deformationsfeld für 2 Bilder

# Tomographie



**Abbildung:** Vier radiographische Projektionen eines 3D-Objektes für seine tomographische Rekonstruktion

## hidden entity



**Abbildung:** (a) Bild, das ein alphanumerisches Zeichen zu enthalten scheint; (b)-(c) Bild c scheint zutreffender zu sein als Bild b; (d) Bayes'sche Schätzung

# Rastergrafiken/Bitmap

- Methode zur Beschreibung zweidimensionaler Bilder in Form von Daten
- kontinuierliche Szene (Objekte, Farbverläufe etc.) wird diskretisiert
- ein (Elementar-)Bereich der Szene wird auf je ein Pixel (Picture Elements) der Bildmatrix abgebildet
- Formate: JPG, BMP, GIF, PNG, TIFF, PSD

# Vektorgrafiken

- Computerbild, das aus grafischen Primitiven wie Linien, Kreisen und Polygonen besteht
- kann ohne Qualitätsverlust stufenlos skaliert und verzerrt werden
- Formate: CDR, DXF (AutoCAD)

# Metagrafikformate

- können Bildinformationen als Rastergrafiken und Vektorgrafiken gemeinsam speichern
- eignen sich gut für den Datenaustausch
- Formate: EPS (Encapsulated PostScript)

# Binärbild

- Pixel können nur die zwei Farben Schwarz oder Weiß annehmen
- Kodierung der Pixel erfolgt meist mit dem Wert 0 für Schwarz und 1 für Weiß  
(bei manchen Modellen -1,1 sinnvoller)
- Speicherbedarf pro Pixel beträgt 1 Bit

# Graustufenbild

- Pixel können  $2^n$  Graustufen annehmen (meist 256)
- Speicherbedarf pro Pixel beträgt  $n$  Bit



# Farbbild

verschiedene Farbmodelle

**RGB** additives Farbmodell aus den Grundfarben **R**ed, **G**reen, **B**lue

- 3 Farbkanälen mit je 8 Bit
- 16.777.216 ( $256^3$ ) verschiedene Farben
- Speicherbedarf pro Pixel beträgt 24 Bit

**CMYK** subtraktives Farbmodell aus den Grundfarben **C**yan, **M**agenta, **Y**ellow, **K**ey

# bedingte Wahrscheinlichkeit

- $(\Omega, \mathcal{A}, \mathcal{P})$  beliebiger Wahrscheinlichkeitsraum
- $B$  beliebiges Ereignis mit  $P(B) > 0$
- bedingte Wahrscheinlichkeit von  $A$  unter  $B$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Satz von Bayes

- $(\Omega, \mathcal{A}, \mathcal{P})$  beliebiger Wahrscheinlichkeitsraum
- $B_1, B_2, \dots$  Zerlegung von  $\Omega$ ,  $P(B_i) > 0$
- dann gilt

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_k P(B_k)P(A|B_k)} \quad \forall i$$

- $P(B_i)$  a-priori-Wahrscheinlichkeit
- $P(B_i|A)$  a-posteriori-Wahrscheinlichkeit

## Idee

- Rekonstruktion eines i.A. unbekanntes Idealbildes  $x$  aus Datensatz  $y$  (gestörte Version von  $x$ )
- inverses Problem (praktisch nie eindeutig lösbar)
- Zusatzbedingungen um Lösungsmenge einzuschränken, z.B. Gütefunktion  $\Pi(x)$
- $P(y|x)$  Wahrscheinlichkeit  $y$  zu beobachten wenn  $x$  das richtige Bild ist (Störung)

# a priori Verteilung

- $X$  Raum aller Bilder
- $\Pi : X \rightarrow \mathbb{R}^+$  Gütefunktion und Wahrscheinlichkeitsverteilung
- $\Pi$  modelliert Vorwissen über die Parameter
- $\Pi(x) > \Pi(x')$  bedeutet:  
   $x$  hat erwünschtere Eigenschaften als  $x'$
- $\Pi$  hängt nur vom Bild  $x$  ab, nicht von den Daten  $y$
- Problem: Konstruktion 'passender' Priors

# Gibbs'sche Form

- Gibbs'sche Form

$$\Pi(x) = Z^{-1} \exp(-K(x)), Z = \sum_z \exp(-K(z))$$

mit  $K : X \rightarrow \mathbb{R}, x \mapsto K(x)$

- $K$  heißt Energie-Funktion
- keine Einschränkung
- jede strikt positive Wahrscheinlichkeitsverteilung  $\Pi$  genügt

$$\Pi(x) = \exp(-K(x)), Z = \sum_z \exp(-K(z)) = \sum_z \Pi(z) = 1$$

(mit  $K(x) = -\ln \Pi(x)$ )

# a posteriori Verteilung

- nach Beobachtung von  $y$  wird  $\Pi$  zu  $\Pi(\cdot|y)$  modifiziert
- $\Pi(x|y) = \frac{P(y|x)\Pi(x)}{\Pi(y)} \propto P(y|x)\Pi(x)$   
( $\Pi(y)$  Randverteilung, hängt nur von  $y$  ab)
- Ziel: mit Hilfe der a posteriori Verteilung möglichst gute Rekonstruktionen  $x$  aus den Daten  $y$  abzuleiten

# Idee

- 'bestes' Bild aus allen 'erlaubten' auswählen
- 'gute' Schätzung muss Daten und Prior 'entsprechen'
- Ausgewogenheit von 'Datentreue' und a-priori-Erwartung
- Bayes'sches Paradigma vereinbart beide Anforderungen

aus *G. Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer, 2nd ed., 2003.*



# Schätzer

- Abbildung  $\hat{X} : Y \rightarrow X, y \mapsto \hat{X}(y)$
- 'Genauigkeit' über Verlustfunktion definiert
- Verlustfunktion  $L : X \times X \rightarrow [0, \infty)$  erfüllt  $L(x, x) = 0$
- Beispiel:

$$L(x, \hat{x}) = \begin{cases} 0 & \text{falls } x = \hat{x} \\ 1 & \text{falls } x \neq \hat{x} \end{cases}$$

(diskrete Verlustfunktion bzw. Abstandsfunktion)

# Bayesrisiko

- Parameter  $x$  durch a priori Verteilung gewichtet
- Bayesrisiko eines Schätzers  $\hat{X}(y)$  für  $x$

$$R(\hat{X}) = \sum_{x,y} L(x, \hat{X}(y)) \Pi(x) P(y|x)$$

# Bayesschätzer

- Schätzer der das Bayesrisiko minimiert
- Berechnung:  $R(\hat{X})$  minimal  $\Leftrightarrow$

$$\mathbb{E}(\hat{X}|y) = \sum_x L(x, \hat{X}(y))\Pi(x|y)$$

(posterior expectation)

## MAP

- maximum posterior estimate  
(Maximum-A-Posteriori-Schätzung)
- maximiert  $\Pi(\cdot|y)$
- oft zu 'global'
- schwer zu berechnen (Integration meist nicht in geschlossener Form möglich)
- beruht auf o.g. Eins-Null-Verlustfunktion

$$L(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

# MMSE

- posterior mean squares estimate
- $\bar{x} = \sum x \Pi(x|y)$
- leichter zu berechnen als MAP
- beruht auf quadratischer Verlustfunktion

$$L(x, \hat{x}) = \sum_s (x_s - \hat{x}_s)^2$$

## MPME

- marginal posterior mode estimate
- $\hat{x}$  MPME  $\Leftrightarrow \hat{x}_s$  Modus von  $x_s \mapsto \Pi(x_s|y)$
- beruht auf error-rate,  $S$  Indexmenge der Pixel

$$L(x, \hat{x}) = |S|^{-1} |s \in S : x_s \neq \hat{x}_s| = |S|^{-1} d(x, \hat{x})$$

- Hamming-Abstand

$$d(x, \hat{x}) = |s \in S : x_s \neq \hat{x}_s|$$

# ICM

- iterated conditional mode
- local maximum of the posterior distribution 'next' to data
- mathematisch schwer zu analysieren
- häufig verwendet
- keine Verlustfunktion bekannt

# stochastischer Prozess

- $(\Omega, \mathcal{A}, \mathcal{P})$  Wahrscheinlichkeitsraum
- $T \neq \emptyset$  Indexmenge
- $(I, \mathcal{I})$  messbarer Raum (Zustandsraum)
- Familie  $(X_t)_{t \in T}$  von Zufallsvariablen mit Werten in  $I$  heißt stochastischer Prozess



# Markov-Kette

- $(X_n)_{n \in \mathbb{Z}^+}$  zufällige Folge
- $I$  abzählbar
- $(X_n)$  Markov-Kette  $\Leftrightarrow$

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n)$$

$$\forall n \in \mathbb{Z}^+ \forall i_0, \dots, i_{n+1} \in I : P(X_0 = i_0, \dots, X_n = i_n) > 0$$

# Eigenschaften I

- Markovsche Kette  $X_0, X_1, \dots$  homogen (Kette mit stationären Übergangswahrscheinlichkeiten), wenn

$$P(X_{n+1} = j | X_n = i) = p_{ij} \quad \forall i, j \in I$$

unabhängig von  $n$  ist

- $P = (p_{ij})$  stochastische Matrix, d.h.

$$p_{ij} \geq 0 (i, j \in I) \quad \text{und} \quad \sum_{j \in I} p_{ij} = 1 (i \in I)$$

# Eigenschaften II

- $\alpha = (\alpha_i, i \in I)$  mit  $\alpha_i = P(X_0 = i)$  heißt Anfangsverteilung
- Anfangsverteilung  $\alpha$  stationär, falls  $\alpha^T = \alpha^T P$
- $p_{ij}^{(n)} := P(X_n = j | X_0 = i)$  Wahrscheinlichkeit des Überganges vom Zustand  $i$  in den Zustand  $j$  in  $n$  Schritten
- $P^{(n)} = (p_{ij}^{(n)})$   $n$ -stufige-Übergangsmatrix

# Ergodizität

Markov-Kette  $X_0, X_1, \dots$  mit Übergangsmatrix  $P = (p_{ij})$  bzw.  $n$ -stufigen Übergangsmatrizen  $P^{(n)} = (p_{ij}^{(n)})$  heißt ergodisch, falls die Grenzwerte

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

- für jedes  $j \in I$  existieren,
- positiv sind und nicht von  $i \in I$  abhängen,
- eine Wahrscheinlichkeitsfunktion  $\pi = (\pi_1, \dots, \pi_I)^\top$  bilden, d.h.,  $\sum_{j \in I} \pi_j = 1$ .

# Irreduzibilität

- Zustand  $j$  vom Zustand  $i$  erreichbar, falls  $p_{ij}^{(n)} > 0$  für ein  $n \geq 0$  gilt ( $i \rightarrow j$ )
- Erreichbarkeit ist transitiv ( $i \rightarrow k$  und  $k \rightarrow j \Rightarrow i \rightarrow j$ )
- Zustände  $i$  und  $j$  kommunizieren, falls  $i \rightarrow j$  und  $j \rightarrow i$  ( $i \leftrightarrow j$ )
- $\leftrightarrow$  ist Äquivalenzrelation
- Markov-Kette  $X_n$  irreduzibel, wenn  $I$  nur aus einer solchen Äquivalenzklasse besteht ( $i \leftrightarrow j \quad \forall i, j \in I$ )
- kurz:  $\forall i, j \in \{1, \dots, k\} \exists n \in \mathbb{N} : p_{ij}^{(n)} > 0$

# Aperiodizität

- Periode  $d_i$  des Zustandes  $i \in I$  ist gegeben durch
$$d_i = \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\}$$
- $d_i = \infty$ , falls  $p_{ii}^{(n)} = 0$  für jedes  $n \geq 1$
- Zustand  $i \in I$  heißt aperiodisch, falls  $d_i = 1$
- Markov-Kette  $\{X_n\}$  heißt aperiodisch, wenn sämtliche Zustände von  $\{X_n\}$  aperiodisch sind
- kurz:  $\forall i \in \{1, \dots, k\} : \text{ggT}\{n \geq 1 : p_{ii}^{(n)} > 0\} = 1$

# Existenz und Eindeutigkeit der stationären Verteilung

## Satz

Jede irreduzible, aperiodische Markov-Kette hat genau eine stationäre Verteilung.

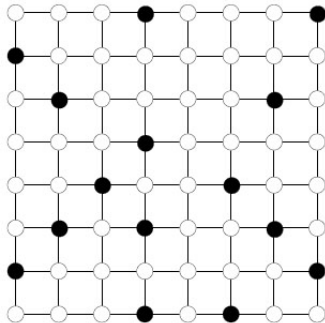
# Hard-Core-Modell

- $G = (V, K)$  Graph
- $V = \{v_1, \dots, v_{|V|}\}$  (endlich viele) Eckpunkte
- $K \subset V^2$  Kanten (verbinden jeweils zwei Eckpunkte)
- jeder Eckpunkt aus  $V$  hat Wert 0 oder 1, so dass keine 2 verbundenen Ecken den Wert 1 haben
- $E \subset \{0, 1\}^{|V|}$  zulässige Konfigurationen
- wähle jedes  $x \in E$  mit gleicher Wahrscheinlichkeit

$$p_x = \frac{1}{|E|} \quad \forall x \in E$$



## Hard-Core-Modell (2)



**Abbildung:** Quadratisches Gitter  $G$  der Größe  $8 \times 8$ , schwarz gesetzte Pixel entsprechen dem Wert 1

# Hard-Core-Modell (3)

- wenn  $|V|$  und  $|K|$  groß sind ist Bestimmung der zulässigen Konfigurationen  $E$  schwierig
- Anzahl  $|E|$  aller zulässigen Konfigurationen typischerweise unbekannt
- $n(x)$  Anzahl Einsen von  $x \in E$
- bestimme erwartete Anzahl von Einsen in  $x'$

$$\mathbb{E}(n(x')) = \sum_{x \in E} n(x) p_x = \frac{1}{|E|} \sum_{x \in E} n(x) = ?$$

# Hard-Core-Modell (4)

- Lösung: konstruiere Markov-Kette  $X_0, X_1, \dots$  mit Zustandsraum  $E$  und ergodischer (Grenz-) Verteilung  $\rho_x$
- erzeuge Pfad  $x_0, x_1, \dots$  der Markov-Kette
  - 1 wähle zulässige Anfangskonfiguration  $x_0 \in E$
  - 2 wähle Ecke  $v \in V$  beliebig und werfe faire Münze
  - 3  $x_{n+1}(v) = 1$  falls  $x_n(w) = 0$  für alle mit  $v$  verbundenen Ecken  $w \in V$  und 'Kopf', 0 sonst
  - 4  $x_{n+1}(w) = x_n(w)$  für jedes  $w \neq v$

## Idee

- konstruiere irreduzible, aperiodische Markov-Kette  $X_n$  mit gewünschter ergodischer Grenzverteilung  $\Pi$
- homogene Markov-Kette ist durch Anfangsverteilung  $\alpha$  und Übergangswahrscheinlichkeiten  $P$  festgelegt
- höherdimensionale Probleme vereinfachen durch sukzessive Simulation von niedriger-dimensionalen Problemen  
⇒ Gibbs-Sampler
- zufällige Markov-Kette korrigieren um stationäre Verteilung  $\Pi$  zu erhalten  
⇒ Metropolis-Hastings-Algorithmus

# Gibbs-Sampler

- $V$  eine endliche (nichtleere) Index-Menge
- $E \subset \mathbb{R}^{|V|}$  endliche (Zustands-) Menge
- $\mathbf{X} = (X(v), v \in V)$  diskreter Zufallsvektor der mit Wahrscheinlichkeit 1 seine Werte in  $E$  annimmt
- $\pi = (\pi_{\mathbf{x}}, \mathbf{x} \in E)$  Wahrscheinlichkeitsfunktion von  $\mathbf{X}$  mit  $\pi_{\mathbf{x}} > 0 \quad \forall \mathbf{x} \in E$

- 

$$\pi_{x(v)|\mathbf{x}(-v)} = P(X(v) = x(v) \mid \mathbf{X}(-v) = \mathbf{x}(-v))$$

mit  $\mathbf{X}(-v) = (X(w), w \in V \setminus \{v\})$

# Gibbs-Sampler (2)

## Algorithmus

- 1 wähle Anfangszustand  $x_0 \in E$
- 2 wähle eine Komponente  $v \in V$  gemäß einer (vorgegebenen) Wahrscheinlichkeitsfunktion  $q = (q_v, v \in V)$
- 3 bestimme aktualisierten Wert  $x_{n+1}(v)$  der  $v$ -ten Komponente gemäß

$$\pi_{\cdot | \mathbf{x}_n(-v)} = (\pi_{x(v) | \mathbf{x}_n(-v)}, \forall x(v) \text{ mit } (x(v), \mathbf{x}(-v)) \in E)$$

- 4  $x_{n+1}(w) = x_n(w)$  für jedes  $w \neq v$

# Metropolis-Hastings-Algorithmus

- Gibbs-Sampler Spezialfall vom Metropolis-Hastings-Algorithmus
- $q(x, y)$  Übergangswahrscheinlichkeit von  $x$  nach  $y$
- $p(y)$  Wahrscheinlichkeit, dass Zustand  $y$  der neue vorgeschlagene Zustand ist
- $p(x)$  Wahrscheinlichkeit, dass Zustand  $x$  der aktuelle Zustand ist
- Akzeptanz-Wahrscheinlichkeit  $\alpha(x, y) = \frac{p(y)q(y, x)}{p(x)q(x, y)}$  (Metropolis-Hastings-Quotient) entscheidet, ob und mit welcher Wahrscheinlichkeit Übergang von einem Zustand in den Anderen akzeptiert wird
- suche einen 'Kandidaten'  $y^*$  und akzeptiere/verwerfe nach bestimmten Kriterien

# Metropolis-Hastings-Algorithmus (2)

## Algorithmus

- 1 gegeben  $X_n = x$ , bestimme  $y^*$  mit  $q(x, \cdot)$
- 2 berechne  $\alpha(x, y^*)$
- 3 akzeptiere  $y^*$  mit Wahrscheinlichkeit  $\alpha(x, y^*) \Rightarrow X_{n+1} = y^*$ ,  
sonst  $X_{n+1} = x$
- 4 wiederhole



# Literatur

- S. Banerjee, B. P. Carlin. Hierarchical Modelling and Analysis for Spatial Data. Chapman & Hall, 2004.
- G. Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer, 2nd ed., 2003.
- B. Chalmond. Modelling and Inverse Problems in Image Analysis. Springer, 2003.
- J. Moeller, Ed. Spatial Statistics and Computational Methods. Springer, LNS 173. 2003.
- V. Schmidt, Markov-Ketten und Monte-Carlo-Simulation, Skript SS03.
- U. Krengel. Einführung in die Wahrscheinlichkeitstheorie und Statistik. vieweg, 2003.
- <http://de.wikipedia.org>