Introduction to statistical Image Analysis Using Markov Random Fields

Hsin-Cheng Huang

Institute of Statistical Science
Academia Sinica

hchuang@stat.sinica.edu.tw
http://www.stat.sinica.edu.tw/~hchuang/

Markov Chains

Suppose that $x_1, \ldots, x_n$ is a Markov chain

- Markov property

$$p(x_{k+1} \,|\, x_1, \ldots, x_k) = p(x_{k+1}|x_k); \quad k = 1, \ldots, n-1,$$

  or

$$x_{k+1} \perp \{x_1, \ldots, x_{k-1}\} \,\big|\, x_k; \quad k = 1, \ldots, n-1$$

  - Markov random fields
  - Graphical Markov models over undirected graphs

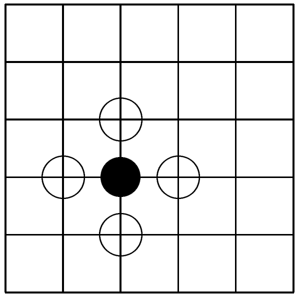
- Joint distribution

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=1}^{n-1} p(x_{i+1}|x_i)$$

  - Partially-ordered Markov models (Markov mesh models)
  - Graphical Markov models over directed graphs

Graphical Markov Models

- A *graph* G is made up of the pair (V, E), where V denotes a finite set of nodes, and $E \subset V \times V$ denotes the set of edges between two distinct nodes



- An edge $(i, j) \in E$ such that $(j, i) \in E$ is called an *undirected edge*, where j is called a *neighbor* of i, and is denoted by $i \sim j$
- $G = (V, E)$ is called an *undirected graph* if E contains only undirected edges
- The *neighborhood set* of i is defined to be

$$\partial i \equiv \{j : j \sim i\}$$

- A *clique* is a subset of V that consists of a single element or of elements that are all neighbors of each other

Markov Properties on Undirected Graphs

Suppose that $G = (V, E)$ is an undirected graph, and $\{x_i : i \in V\}$ is a collection of random variables

- *Pairwise Markov property*

  $$i \perp j \,|\, V \setminus \{i, j\},$$

  for all $i \not\sim j$

- *Local Markov property*

  $$i \perp j \,|\, \partial_i,$$

  for all $j \in V \setminus (\partial_i \cup \{i\})$

- *Global Markov property*

  $$A \perp B \,|\, S,$$

  for all disjoint subsets A, B, S of V such that S separates A from B in G

- Global Markov property $\Rightarrow$ local Markov property $\Rightarrow$ pairwise Markov property

Markov Random Fields (MRFs)

- Specify the joint distribution $p(\boldsymbol{x}) \equiv p(x_1, \ldots, x_n)$ at the local level via the following conditional distributions:

$$p(x_i|\boldsymbol{x}_{-i}) = p(x_i|\boldsymbol{x}_{\partial_i}); \quad i = 1, \ldots, n,$$

  where $x_{-i} = \{x_j : j \neq i\}$, and site $j \in \partial_i \Leftrightarrow p(x_i|\boldsymbol{x}_{-i})$ is depend on x_j

- The joint distribution $p(\boldsymbol{x})$ satisfies the local Markov property with respect to $G = (V, E)$, where $V = \{1, \ldots, n\}$ and $E = \{(i, j) : j \in \partial_i,\ i, j = 1, \ldots, n\}$

- Assume that $p(\boldsymbol{x}) > 0$. Then $p(\boldsymbol{x})$ is determined by its local conditional distributions $\{p(x_i|\boldsymbol{x}_{-i}) : i = 1, \ldots, n\}$

$$p(\boldsymbol{x}) = p(\mathbf{0}) \prod_{i=1}^{n} \frac{p(x_i \,|\, 0, \ldots, 0, x_{i+1}, \ldots, x_n)}{p(0 \,|\, 0, \ldots, 0, x_{i+1}, \ldots, x_n)}$$

- The specification must be symmetric in that $i \in \partial_j \Leftrightarrow j \in \partial_i$, and is usually based on proximity of locations

- The specification must satisfy some consistency conditions needed for reconstruction of a joint distribution

Gibbs Fields

Let $\{x_1, \ldots, x_n\}$ be a collection of random variables

- Perturb i.i.d. fields by local interactions of 2 bodies, of 3 bodies, etc
- Gibbs fields
$$p(\boldsymbol{x}) \propto \exp\left\{-\sum_{C \in \mathcal{C}} V_C(\boldsymbol{x}_C)\right\},$$
where $\mathcal{C} \subset 2^{\{1,\ldots,n\}}$, and $\int \exp\left\{-\sum_{C \in \mathcal{C}} V_C(\boldsymbol{x}_C)\right\} d\boldsymbol{x} < \infty$
- $\{V_C\}$ are called *potential functions*. Lower potentials correspond to higher probabilities and vice versa
- The full conditionals are of the form
$$p(x_i|\boldsymbol{x}_{-i}) \propto \exp\left\{-\sum_{C: i \in \mathcal{C}} V_C(\boldsymbol{x}_C)\right\}$$

Hammersley and Clifford Theorem (1971)

Consider an undirected graph $G = (V, E)$, and a collection of random variables $\boldsymbol{x} \equiv \{x_i : i \in V\}$. Suppose that the joint distribution of $\boldsymbol{x}$ has positive and continuous density $p(\boldsymbol{x})$ with respect to a product measure. Then the following statements are equivalent:

- $p(\boldsymbol{x})$ satisfies the pairwise Markov property with respect to G
- $p(\boldsymbol{x})$ satisfies the local Markov property with respect to G
- $p(\boldsymbol{x})$ satisfies the global Markov property with respect to G
- $p(\boldsymbol{x})$ can be written as:

$$p(\boldsymbol{x}) \propto \exp\left\{-\sum_{C \in \mathcal{C}} V_C(\boldsymbol{x}_C)\right\},$$

where $\mathcal{C}$ is the set of cliques of G

Besag's Auto Models (Besag, 1974)

- Assumption 1: pairwise-only dependence

$$p(\boldsymbol{x}) \propto \exp\left\{-\sum_{i=1}^{n} V_i(x_i) - \sum_{1\le i<j\le n} V_{ij}(x_i, x_j)\right\}$$

- Assumption 2: conditional exponential distributions

$$p(x_i|\partial_i) = \exp\left\{A_i(\partial_i)B_i(x_i) + C_i(x_i) + D_i(\partial_i)\right\}$$

- Auto models

$$\begin{aligned} p(\boldsymbol{x}) &\propto \exp\left\{\sum_{i=1}^{n}(\alpha_i B_i(x_i) + C_i(x_i)) + \sum_{1\le i<j\le n} \beta_{ij} B_i(x_i) B_j(x_j)\right\}, \\ p(x_i|\partial_i) &= \exp\left\{\alpha_i B_i(x_i) + \sum_{j\in\partial_i} \beta_{ij} B_i(x_i) B_j(x_j) + C_i(x_i) + D_i(\partial_i)\right\} \end{aligned}$$

Conditionally Specified Spatial Models for Discrete Data

- Auto-logistic model

$$p(\boldsymbol{x}) \;\propto\; \exp\left\{\sum_{i=1}^{n}\alpha_i x_i + \sum_{1\le i<j\le n}\beta_{ij}x_i x_j\right\},$$
$$p(x_i|\boldsymbol{x}_{-i}) \;=\; \frac{\exp\left\{x_i\left(\alpha_i + \Sigma_{j\in\partial_i}\beta_{ij}x_j\right)\right\}}{1+\exp\left(\alpha_i + \Sigma_{j\in\partial_i}\beta_{ij}x_j\right)}$$

  - Ferromagnetism in statistical mechanics (Ising model)
  - Presence/absence of plants
  - Black-and-white images
  - A limiting Markov random field exists, but may not be unique (phase transition)

- Auto-binomial model

$$p(\boldsymbol{x}) \;\propto\; \exp\left\{\sum_{i=1}^{n}\alpha_i x_i + \sum_{1\le i<j\le n}\beta_{ij}x_i x_j + \sum_{i=1}^{n}\log\left(\binom{n_i}{x_i}\right)\right\},$$
$$p(x_i|\boldsymbol{x}_{-i}) \;=\; \binom{n_i}{x_i}\left(\frac{\exp(\alpha_i + \Sigma_{j\in\partial_i}\beta_{ij}x_j)}{1+\exp(\alpha_i + \Sigma_{j\in\partial_i}\beta_{ij}x_j)}\right)^{x_i}\left(\frac{1}{1+\exp(\alpha_i + \Sigma_{j\in\partial_i}\beta_{ij}x_j)}\right)^{n_i-x_i}$$

Conditionally Specified Spatial Models for Discrete Data — Cont.

- Auto-Poisson model

$$
\begin{aligned}
p(\boldsymbol{x}) &\propto \exp\left\{ \sum_{i=1}^{n} \alpha_i x_i + \sum_{1 \le i < j \le n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \log(x_i!) \right\}, \\
p(x_i | \boldsymbol{x}_{-i}) &= \frac{\exp\left\{- \exp(\alpha_i + \Sigma_{j \in \partial_i} \beta_{ij} x_j)\right\} \left\{\exp(\alpha_i + \Sigma_{j \in \partial_i} \beta_{ij} x_j)\right\}^{x_i}}{x_i!}
\end{aligned}
$$

  - Competitive interaction (negative spatial dependence): $\beta_{ij} \le 0$
  - Positive spatial dependence can be obtained by using Winsorized Poisson conditionals (Kaiser and Cressie, 1997)

- Auto negative binomial model

Conditionally Specified Spatial Models for Continuous Data

- Auto-Gaussian model

$$
\begin{aligned}
p(\boldsymbol{x}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{M}^{-1}(\boldsymbol{I}-\boldsymbol{C})(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \\
p(x_i|\boldsymbol{x}_{-i}) &= \left(2\pi\tau_i^2\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\tau_i^2}\left(x_i-\mu_i-\sum_{j\in\partial_i} c_{ij}(x_j-\mu_j)\right)^2\right\},
\end{aligned}
$$

  where $\boldsymbol{C}$ is an $n \times n$ matrix whose (i,j)-th element is c_{ij}, $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$, $\boldsymbol{M} = diag(\tau_1^2, \ldots, \tau_n^2)$, and $\mu_i = E(x_i)$

- Auto-Gamma model

Infinite Lattice

Let $\mathcal{S}$ be the collection of nonempty finite subsets of $\mathbb{Z}^d$

- For a coherent family of marginal distributions $\{P_V : V \in \mathcal{S}\}$, there exists a unique field μ such that $\mu_V = P_V$; $V \in \mathcal{S}$
- For a coherent family of conditional probability kernel $\{\pi_V : V \in \mathcal{S}\}$, does there exist a field μ such that its conditional distributions are the ones considered?
- Sufficient condition for the existence and the uniqueness of a Gibbs field is given by Dobrushin (1968)

Edge-Effect Correction

- Conditional likelihood by conditioning on the edge sites
- Wrap the n by m lattice on a torus
- Assume some values for unobserved neighbors
- Integrate out the presence of the unobserved neighbors

Simulation

- Exact simulation
  - Auto-normal
  - Auto-logistic
  - Markov mesh models
  - Partially ordered Markov models
- Rejection sampling
- Markov chain Monte Carlo
  - Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970)
  - Gibbs sampler (Geman and Geman, 1984)

Applications

- Remote sensing by satellites
  - Remote-sensing data may be used for the estimation of crop yields, classification of pixels into land uses or crop types, locating ice floes, or for the monitoring of desertification, forests, erosion, geological mapping, and pollution
  - Statistical techniques are required for the processing and classification of such data
  - Conventional methods of multivariate analysis (e.g., linear discriminant analysis, principal components analysis, and cluster analysis) do not really exploit the spatial component of the data, and so may result in a high error rate

Applications — Cont.

- Medical images
  - CAT scan
  - Ultra sound
  - Magnetic resonance imaging (MRI)
  - Functional MRI (fMRI)
  - Positron emission tomography (PET)
- Optical astronomy radar
- Electron microscope imaging
- Photography
- Computer vision and graphics
- Object recognition

Some Problems in Image Analysis

- Image restoration
  - Recovering of the true image from the contaminated image data
- Boundary detection
  - Boundary detection is to find locations in the image which correspond to physical discontinuous in the actual scene
- Image segmentation
  - Segmentation refers to partitioning of the domain into homogeneous (possibly unlabeled) regions based on local properties such as color, depth, surface orientation, or texture
  - Texture refers to pattern or regularity in the local spatial distribution of gray levels, like sand, grass, wood, or damaged wood. It is often the dominant feature in satellite, medical and other data
- Image classification
  - Assign labels to pixels or regions

Some Problems in Image Analysis — Cont.

- Texture synthesis
  - Texture can be simulated by realizations of stochastic processes
  - Auto-binomial MRF (Besag, 1974)
  - Markov mesh random fields (Abend *et al.*, 1965)
  - Partially ordered Markov random fields (Cressie and Davidson, 1998)

Distortion and Artifacts

- Blur
  - Scattering within the medium (e.g., atmosphere, human body)
  - De-focused camera lenses
  - Motion of the median, objects, or camera
- Noise
  - Random nature of photo propagation
  - Sensing and recoding devices
- Nonlinear transformation
  - Sensing and recoding devices

Image Problem Formulation

Estimation of the image $\boldsymbol{x} \equiv (x_1, \ldots, x_n)'$ from the contaminated-image data $\boldsymbol{y} \equiv (y_1, \ldots, y_n)'$

$$\boldsymbol{y} = \phi(\boldsymbol{H}\boldsymbol{x}) \odot \boldsymbol{\epsilon}$$

- $\boldsymbol{H}$ is a blurring matrix (point spread function), which is usually assumed to be translation-invariant
- ϕ is an increasing function accounting for sensor effects
  - Photoelectric sensors: ϕ is a linear transformation
  - Photochemical sensors: ϕ is often a known nonlinear function
- $\boldsymbol{\epsilon}$ is a noise process
- Example

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

Bayesian Image Analysis

- Prior distribution of $\boldsymbol{x} \in \Omega$:
$$\pi(\boldsymbol{x})$$
- Conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$:
$$f(\boldsymbol{y}|\boldsymbol{x})$$
- Posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$ is
$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})}{\int f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x}}$$
- Assume $\pi(\boldsymbol{x})$ is a MRF over graph G, and
$$f(\boldsymbol{y}|\boldsymbol{x}) = \sum_{i=1}^{n} g_i(y_i|x_i).$$
Then $p(\boldsymbol{x}|\boldsymbol{y})$ is also a MRF over G
- Compromise of both satisfying the prior distribution $\pi(\boldsymbol{x})$ and being faithful to observations $\boldsymbol{y}$

Bayesian Estimators

- *Maximum a posteriori (MAP) estimator*
  - Bayes estimate for the loss function

$$L(\boldsymbol{x}, \hat{\boldsymbol{x}}) = I(\boldsymbol{x} \neq \hat{\boldsymbol{x}})$$

  - Iterated conditional modes

$$\arg\sup_{x_1} p(x_1|\boldsymbol{y}, \boldsymbol{x}_{-1}), \ldots, \arg\sup_{x_p} p(x_p|\boldsymbol{y}, \boldsymbol{x}_{-p})$$

  - Simulated annealing

$$p_T(\boldsymbol{x}|\boldsymbol{y}) = \frac{\left(f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})\right)^{1/T}}{\int \left(f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})\right)^{1/T} d\boldsymbol{x}}$$

- *Minimum mean-squared error estimator*
  - Bayes estimate for the loss function

$$L(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \sum_{i=1}^{n} \left(x_i - \hat{x}_i\right)^2$$

Bayesian Estimators — Cont.

- *Maximum posterior marginal probability estimator*

$$\hat{\boldsymbol{x}} = \left(\arg\sup_{x_1} p(x_1|\boldsymbol{y}), \ldots, \arg\sup_{x_p} p(x_p|\boldsymbol{y}) \right)'$$

  – Bayes estimate for the loss function

$$L(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \sum_{i=1}^{n} I(x_i = \hat{x}_i),$$

  i.e., the misclassification rate

Simulated Annealing

Simulated annealing is a sequential procedure made up of a Gibbs sampler and an annealing schedule

- Gibbs sampler

$$p_T(\boldsymbol{x}|\boldsymbol{y}) = \frac{\left(f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})\right)^{1/T}}{\int \left(f(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x})\right)^{1/T} d\boldsymbol{x}}; \quad T > 0$$

$$\begin{aligned} T = 1: &\quad \text{posterior distribution;} \\ T \to \infty: &\quad p_T(\cdot|\boldsymbol{y}) \text{ approaches a uniform distribution on } x; \\ T \to 0: &\quad p_T(\cdot|\boldsymbol{y}) \text{ becomes concentrated on the MAP estimate } \hat{\boldsymbol{x}} \end{aligned}$$

- Annealing schedule
  - The rate at which $T(t) \to 0$
  - Geman and Geman (1984) show that if

$$\liminf_{t\to\infty} T(t) \log t = r,$$

for some r sufficient large. Then $\hat{\boldsymbol{x}}_t \to \hat{\boldsymbol{x}}$

Markov-Random-Field Models in Image Analysis

- Binary image with $\Omega = \{0,1\}^n$ (Creig *et al.*, 1986, 1989)

$$\pi(\boldsymbol{x}) \propto \exp\left\{\frac{1}{2}\sum_{1\le i<j\le n}\beta_{ij}\big(x_i x_j + (1-x_i)(1-x_j)\big)\right\}; \quad \beta_{ij} > 0,$$

i.e., β_{ij} is included in the summation iff $x_i = x_j$

- Polychotomous image with $\Omega = \{1,2,\ldots,K\}^n$ (Besag, 1986)

$$\pi(\boldsymbol{x}) \propto \exp\left\{\sum_{k=1}^{K}\alpha_k n_k - \sum_{k\ne l}\beta_{kl}n_{kl}\right\},$$

where

$$\begin{aligned} n_k &\equiv \#\text{ pixels colored } k, \\ n_{kl} &\equiv \#\text{ pixels colored } (k,l) \end{aligned}$$

Markov-Random-Field Models in Image Analysis — Cont.

- Edge detection (Geman and Geman, 1984; Chalmond, 1988)
  - Graph $G = (V, E)$, where $V = \{1, \ldots, n\}$ is the set of pixel sites, and E consists of all nearest-neighbor pairs (i, j) from V
  - Edge process
  $$\boldsymbol{x}_E \equiv \{x_{(i,j)} : i \sim j,\ i, j \in V\},$$
  where $x_{(i,j)} = 1$ (resp. 0) indicates presence (resp. absence) of an edge (i, j)
  - The data $\boldsymbol{y}_V$ are observed according to
  $$p(\boldsymbol{y}_V | \boldsymbol{x}_V),$$
  where $\boldsymbol{x}_V \in \{0, 1, \ldots, L\}^n$
  - MRF prior for $\boldsymbol{x} = (\boldsymbol{x}_V, \boldsymbol{x}_E)$
  $$\pi(\boldsymbol{x}) \propto \exp\left\{-H(\boldsymbol{x})\right\},$$
  where
  $$H(\boldsymbol{x}) = H_1(\boldsymbol{x}_V, \boldsymbol{x}_E) + H_2(\boldsymbol{x}_E)$$

Markov-Random-Field Models in Image Analysis — Cont.

- Edge detection (Geman and Geman, 1984; Chalmond, 1988) — Cont.
  - H_1 reflects expectations about interactions between intensities and edges

$$H_1(\boldsymbol{x}_V, \boldsymbol{x}_E) = \theta_1 \sum_{i \sim j} \phi(x_i - x_j)(1 - x_{(i,j)}), \quad \theta_1 > 0,$$

  where $\phi(0) = -1$, and ϕ is even and non-decreasing on $[0, \infty)$
  - H_2 reflects boundary organization

$$H_2(\boldsymbol{x}_E) = -\theta_2 \sum_{k=2}^{6} \xi_k, \quad \theta_2 > 0$$

o	o	o o	o	o	o	o	o o	o o
— —	—	— —	—	— —	
o	o	o o	o o	o o	o o	o o
ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6

  - A simple choice is $\xi_1 = \xi_2 < \xi_3 < \xi_4 = \xi_5 = \xi_6$

Markov-Random-Field Models in Image Analysis — Cont.

- Texture segmentation – boundary detection (Geman *et al.*, 1990)
  - Boundary process
    $$\boldsymbol{y} \equiv \{y_s : s \in S\},$$
    where $y_i = 1$ (resp. 0) indicates presence (resp. absence) of a boundary at i, and S is a regular lattice, distinct from that of the pixels
  - MRF prior for $\boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{y})$ with forbidden patterns
    $$\pi(\boldsymbol{x}, \boldsymbol{y}) \propto \exp\{-H(\boldsymbol{x}, \boldsymbol{y})\}\, I(J(\boldsymbol{y}) = 0),$$
    where $J(\boldsymbol{y})$ is the number of penalties in $\boldsymbol{y}$
  - H is chosen to promote placement of boundaries between regions in the image that demonstrate distinct spatial patterns:
    $$H(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i \sim j} \Phi_{i,j}(\boldsymbol{x})(1 - y_i\, y_j)$$
  - The function $\Phi_{i,j}$ is a measure of disparity between the gray-level values in two blocks of pixels B_i, B_j adjacent to (i, j)
  - Large disparities ($\Phi_{i,j} > 0$) encourage the presence of a boundary ($y_i = y_j = 1$), and vice versa

- Texture segmentation – boundary detection (Geman *et al.*, 1990) — Cont.
  - One possible choice of $\Phi_{i,j}$ using the Kolmogorov-Smirnov distance is

$$\Phi_{i,j} = 2\, I\left(d(y_{B_1}, y_{B_j}) > c\right) - 1$$

  - The penalty function $J(\boldsymbol{y})$ that counts the number of penalties is chosen to inhibit unwanted configurations, such as those defining very small regions, redundant boundary, or blindly ending boundaries, e.g., a penalty for very small regions is

$$\sum_{s \in S} I\left(\sum_{s' \in B_s} I(y_s = y_{s'}) < 9\right),$$

    where B_s is the 5×5 block of sites in S centered at s

Markov-Random-Field Models in Image Analysis — Cont.

- Texture segmentation – partitioning (Geman *et al.*, 1990)
  - Partitioning process
$$\boldsymbol{y} \equiv \{y_s : s \in S\},$$
where $y_i \in \{0, \ldots, K-1\}$ is the label of site i. The maximum number of label is fixed *a priori*
  - MRF prior for $\boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{y})$ with forbidden patterns
$$\pi(\boldsymbol{x}, \boldsymbol{y}) \propto \exp\{-H(\boldsymbol{x}, \boldsymbol{y})\}\, I(J(\boldsymbol{y}) = 0)$$
  - H is chosen to promote placement of boundaries between regions in the image that demonstrate distinct spatial patterns:
$$H(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i \sim j} \Phi_{i,j}(\boldsymbol{x})\, I(y_i = y_j)$$
  - The function $\Phi_{i,j}$ is a measure of disparity between the gray-level values in two blocks of pixels adjacent to (i, j)

Markov-Random-Field Models in Image Analysis — Cont.

- Single photon emission tomography (Geman and McClure, 1985)
  - The digitized isotope intensity is thought to be a realization of a spatial process $\boldsymbol{x} = \{x_i : i = 1, \ldots, n\}$
  - The idea is to use a MRF prior to reflect that nearby pixels are more alike than those far apart, whereas sharp changes in concentration may occur (e.g., across a boundary between two tissue types)
  - MRF prior $\pi(\boldsymbol{x}) = \exp\{-H(\boldsymbol{x})\}$, where

$$H(\boldsymbol{x}) = \beta \left\{ \sum_{<i,j>} \phi(x_i - x_j) + 2^{-1/2} \sum_{[i,j]} \phi(x_i - x_j) \right\},$$

    $< i, j >$ denotes a nearest-neighbor bond, $[i, j]$ denotes a nearest-neighbor-diagonal bond, and

$$\phi(x) = 1 - \frac{2}{1 + (x/\Delta)^2}, \quad x \in [-L, L]$$

Markov-Random-Field Models in Image Analysis — Cont.

- Single Photon emission tomography (Geman and McClure, 1985) — Cont.
  - The data $\boldsymbol{y} = (y_1, \ldots, y_n)'$, with $\boldsymbol{\lambda} \equiv E(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{x}$, are observed according to
  $$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i}}{y_i!} \exp(-\lambda_i),$$
  where the linear operator $\boldsymbol{A}$ incorporates attenuation via the attenuated Radon transform
  - The choice of the smoothing parameter β is critical. For $\beta = 0$, the MAP estimator is just the ML estimator, hence typically too rough
- Three-dimensional Shape reconstruction from shading (Torreǎo, 1989)

Parameter Estimation with Training Pixels

Suppose there are training pixels V for which true values $\{x_i : i \in V\}$ are known. We would like to estimate the parameter $\boldsymbol{\eta}$ from observed $\boldsymbol{x}$

$$\pi(\boldsymbol{x};\boldsymbol{\eta}) = \frac{\exp\left(\sum_{C\in\mathcal{C}} Q_C(\boldsymbol{x}_C;\boldsymbol{\eta})\right)}{\int \exp\left(\sum_{C\in\mathcal{C}} Q_C(\boldsymbol{x}_C;\boldsymbol{\eta})\right) d\boldsymbol{x}}$$

- *Maximum likelihood*
  - Consistent, asymptotically Gaussian, asymptotically efficient
  - The normalizing constant depends on $\boldsymbol{\eta}$
  - Computationally intensive
- *Coding* (Besag, 1974): $\hat{\boldsymbol{\eta}}$ is chosen to maximize the conditional likelihood

  $$\prod_{i\in V_0} p\left(x_i|\partial_i;\boldsymbol{\eta}\right),$$

  where $V = V_0 \cup V_1$, and $\{x_i : i \in V_0\}$ given $\{x_i : i \in V_1\}$ are mutually independent

Parameter Estimation with Training Pixels — Cont.

- *Maximum pseudo likelihood* (Besag, 1975): $\hat{\boldsymbol{\eta}}$ is chosen to maximize

$$\prod_{i=1}^{n} p(x_i | \partial_i; \boldsymbol{\eta})$$

- Monte Carlo maximum likelihood (Geyer and Thompson, 1995)
- Estimating equation

Further Research

- Properties of MRF without the positivity condition
- Construction of new MRF model classes (e.g., higher-order interactions)
- Edge-effect correction
- Simulation
- Parameter estimation

References

Cressie, N. (1993). *Statistics for Spatial Data*, revised edition. Wiely, New York

Geman, D. (1990). Random fields and inverse problems in imaging. *Lecture Notes in Mathematics*, **1427**. Springer-Verlag, New York

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligent*, **6**, 721-741

Geman, D. and Gidas, B. (1991). Image analysis and computer vision. In *Spatial Statistics and Digital Image Analysis*, Panel on spatial statistics and image processing, ed. National Academy Press, Washington D.C., 9-36

Geman, D., Geman, S., Graffigne, C., and Dong, P. (1990). Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligent*, **12**, 609-628

Guyon, X.(1995). *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer-Verlag, New York

Tjelmeland, H. and Besag, J. (1998). Markov random fields with higher-order interactions. *Scandinavian Journal of Statistics*, **25**, 415-433