Markov Random Fields

Umamahesh Srinivas

iPAL Group Meeting

February 25, 2011

Outline

1. Basic graph-theoretic concepts
2. Markov chain
3. Markov random field (MRF)
4. Gauss-Markov random field (GMRF), and applications
5. Other popular MRFs

References

1. Charles Bouman, *Markov random fields and stochastic image models*. Tutorial presented at ICIP 1995
2. Mario Figueiredo, *Bayesian methods and Markov random fields*. Tutorial presented at CVPR 1998

Basic graph-theoretic concepts

- A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a finite collection of nodes (or vertices) $\mathcal{V} = \{n_1, n_2, \ldots, n_N\}$ and set of edges $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- We consider only undirected graphs
- Neighbor: Two nodes $n_i, n_j \in \mathcal{V}$ are neighbors if $(n_i, n_j) \in \mathcal{E}$
- Neighborhood of a node: $\mathcal{N}(n_i) = \{n_j : (n_i, n_j) \in \mathcal{E}\}$
- Neighborhood is a symmetric relation: $n_i \in \mathcal{N}(n_j) \Leftrightarrow n_j \in \mathcal{N}(n_i)$
- Complete graph: $\forall n_i \in \mathcal{V},\ \mathcal{N}(n_i) = \{(n_i, n_j), j = \{1, 2, \ldots, N\} \backslash \{i\}\}$
- Clique: a complete subgraph of $\mathcal{G}$.
- Maximal clique: Clique with maximal number of nodes; cannot add any other node while still retaining complete connectedness.

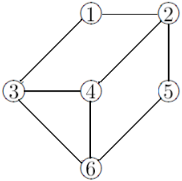
Basic graph-theoretic concepts

- A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a finite collection of nodes (or vertices) $\mathcal{V} = \{n_1, n_2, \ldots, n_N\}$ and set of edges $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- We consider only undirected graphs
- Neighbor: Two nodes $n_i, n_j \in \mathcal{V}$ are neighbors if $(n_i, n_j) \in \mathcal{E}$
- Neighborhood of a node: $\mathcal{N}(n_i) = \{n_j : (n_i, n_j) \in \mathcal{E}\}$
- Neighborhood is a symmetric relation: $n_i \in \mathcal{N}(n_j) \Leftrightarrow n_j \in \mathcal{N}(n_i)$
- Complete graph: $\forall n_i \in \mathcal{V},\ \mathcal{N}(n_i) = \{(n_i, n_j), j = \{1, 2, \ldots, N\} \backslash \{i\}\}$
- Clique: a complete subgraph of $\mathcal{G}$.
- Maximal clique: Clique with maximal number of nodes; cannot add any other node while still retaining complete connectedness.

Basic graph-theoretic concepts

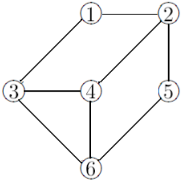
- A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a finite collection of nodes (or vertices) $\mathcal{V} = \{n_1, n_2, \ldots, n_N\}$ and set of edges $\mathcal{E} \subset \binom{\mathcal{V}}{2}$
- We consider only undirected graphs
- Neighbor: Two nodes $n_i, n_j \in \mathcal{V}$ are neighbors if $(n_i, n_j) \in \mathcal{E}$
- Neighborhood of a node: $\mathcal{N}(n_i) = \{n_j : (n_i, n_j) \in \mathcal{E}\}$
- Neighborhood is a symmetric relation: $n_i \in \mathcal{N}(n_j) \Leftrightarrow n_j \in \mathcal{N}(n_i)$
- Complete graph:
  $\forall n_i \in \mathcal{V},\ \mathcal{N}(n_i) = \{(n_i, n_j), j = \{1, 2, \ldots, N\} \backslash \{i\}\}$
- Clique: a complete subgraph of $\mathcal{G}$.
- Maximal clique: Clique with maximal number of nodes; cannot add any other node while still retaining complete connectedness.

Illustration



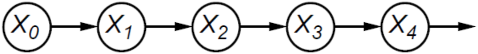
- $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{E} = \{(1, 2), (1, 3), (2, 4), (2, 5), (3, 4), (3, 6), (4, 6), (5, 6)\}$
- $\mathcal{N}(4) = \{2, 3, 6\}$
- Examples of cliques: $\{(1), (3, 4, 6), (2, 5)\}$
- Set of all cliques: $\mathcal{V} \cup \mathcal{E} \cup \{3, 4, 6\}$

Separation



- Let A, B, C be three disjoint subsets of $\mathcal{V}$
- C separates A from B if any path from a node in A to a node in B contains some node in C
- Example: $C = \{1, 4, 6\}$ separates $A = \{3\}$ from $B = \{2, 5\}$

Markov chains



- Graphical model: Associate each node of a graph with a random variable (or a collection thereof)
- Homogeneous 1-D Markov chain:

$$p(x_n|x_i, i < n) = p(x_n|x_{n-1})$$

- Probability of a sequence given by:

$$p(x) = p(x_0) \prod_{n=1}^{N} p(x_n|x_{n-1})$$

2-D Markov chains

$X_{(0,0)}$	$X_{(0,1)}$	$X_{(0,2)}$	$X_{(0,3)}$	$X_{(0,4)}$
$X_{(1,0)}$	$X_{(1,1)}$	$X_{(1,2)}$	$X_{(1,3)}$	$X_{(1,4)}$
$X_{(2,0)}$	$X_{(2,1)}$	$X_{(2,2)}$	$X_{(2,3)}$	$X_{(2,4)}$
$X_{(3,0)}$	$X_{(3,1)}$	$X_{(3,2)}$	$X_{(3,3)}$	$X_{(3,4)}$

- Advantages:
  - Simple expressions for probability
  - Simple parameter estimation
- Disadvantages:
  - No natural ordering of image pixels
  - Anisotropic model behavior

各向异性的

Random fields on graphs

- Consider a collection of random variables $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with associated joint probability distribution $p(\mathbf{x})$
- Let A, B, C be three disjoint subsets of $\mathcal{V}$. Let $\mathbf{x}_A$ denote the collection of random variables in A.
  - Conditional independence: $A \perp\!\!\!\perp B \mid C$
  - $A \perp\!\!\!\perp B \mid C \Leftrightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$
- Markov random field: undirected graphical model in which each node corresponds to a random variable or a collection of random variables, and the edges identify conditional dependencies.

Random fields on graphs

- Consider a collection of random variables $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with associated joint probability distribution $p(\mathbf{x})$
- Let A, B, C be three disjoint subsets of $\mathcal{V}$. Let $\mathbf{x}_A$ denote the collection of random variables in A.
  - Conditional independence: $A \perp\!\!\!\perp B \mid C$
  - $A \perp\!\!\!\perp B \mid C \Leftrightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$
- Markov random field: undirected graphical model in which each node corresponds to a random variable or a collection of random variables, and the edges identify conditional dependencies.

Markov properties

Pairwise Markovianity:

- $(n_i, n_j) \notin \mathcal{E} \Rightarrow x_i$ and x_j are independent when conditioned on all other variables

$$p(x_i, x_j | \mathbf{x}_{\setminus\{i,j\}}) = p(x_i | \mathbf{x}_{\setminus\{i,j\}}) p(x_j | \mathbf{x}_{\setminus\{i,j\}})$$

Local Markovianity:

- Given its neighborhood, a variable is independent on the rest of the variables

$$p(x_i | \mathbf{x}_{\mathcal{V}\setminus\{i\}}) = p(x_i | \mathbf{x}_{\mathcal{N}(i)})$$

Global Markovianity:

- Let A, B, C be three disjoint subsets of $\mathcal{V}$. If

C separates A from $B \Rightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$,

then $p(\cdot)$ is global Markov w.r.t. $\mathcal{G}$.

Markov properties

Pairwise Markovianity:

- $(n_i, n_j) \notin \mathcal{E} \Rightarrow x_i$ and x_j are independent when conditioned on all other variables

$$p(x_i, x_j | \mathbf{x}_{\setminus\{i,j\}}) = p(x_i | \mathbf{x}_{\setminus\{i,j\}}) p(x_j | \mathbf{x}_{\setminus\{i,j\}})$$

Local Markovianity:

- Given its neighborhood, a variable is independent on the rest of the variables

$$p(x_i | \mathbf{x}_{\mathcal{V}\setminus\{i\}}) = p(x_i | \mathbf{x}_{\mathcal{N}(i)})$$ 只与 neighborhood 有关联

Global Markovianity:

- Let A, B, C be three disjoint subsets of $\mathcal{V}$. If

  C separates A from $B \Rightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$,

  then $p(\cdot)$ is global Markov w.r.t. $\mathcal{G}$.

Markov properties

Pairwise Markovianity:

- $(n_i, n_j) \notin \mathcal{E} \Rightarrow x_i$ and x_j are independent when conditioned on all other variables

$$p(x_i, x_j | \mathbf{x}_{\setminus \{i,j\}}) = p(x_i | \mathbf{x}_{\setminus \{i,j\}}) p(x_j | \mathbf{x}_{\setminus \{i,j\}})$$

Local Markovianity:

- Given its neighborhood, a variable is independent on the rest of the variables

$$p(x_i | \mathbf{x}_{\mathcal{V} \setminus \{i\}}) = p(x_i | \mathbf{x}_{\mathcal{N}(i)})$$

Global Markovianity:

- Let A, B, C be three disjoint subsets of $\mathcal{V}$. If

  C separates A from $B \Rightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$,

  then $p(\cdot)$ is global Markov w.r.t. $\mathcal{G}$.

Hammersley-Clifford Theorem

Consider a random field $\mathbf{x}$ on a graph $\mathcal{G}$, such that $p(\mathbf{x}) > 0$. Let $\mathcal{C}$ denote the set of all maximal cliques of the graph.

- If the field has the local Markov property, then $p(\mathbf{x})$ can be written as a Gibbs distribution:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ -\sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) \right\},$$

  where Z, the normalizing constant, is called the partition function; $V_C(\mathbf{x}_C)$ are the clique potentials

- If $p(\mathbf{x})$ can be written in Gibbs form for the cliques of some graph, then it has the global Markov property.

Fundamental consequence: every Markov random field can be specified via clique potentials.

Hammersley-Clifford Theorem

Consider a random field $\mathbf{x}$ on a graph $\mathcal{G}$, such that $p(\mathbf{x}) > 0$. Let $\mathcal{C}$ denote the set of all maximal cliques of the graph.

- If the field has the local Markov property, then $p(\mathbf{x})$ can be written as a Gibbs distribution:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ - \sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) \right\},$$

  where Z, the normalizing constant, is called the partition function; $V_C(\mathbf{x}_C)$ are the clique potentials

- If $p(\mathbf{x})$ can be written in Gibbs form for the cliques of some graph, then it has the global Markov property.

Fundamental consequence: every Markov random field can be specified via clique potentials.

Hammersley-Clifford Theorem

Consider a random field $\mathbf{x}$ on a graph $\mathcal{G}$, such that $p(\mathbf{x}) > 0$. Let $\mathcal{C}$ denote the set of all maximal cliques of the graph.

→ a complete subgraph

- If the field has the local Markov property, then $p(\mathbf{x})$ can be written as a Gibbs distribution:

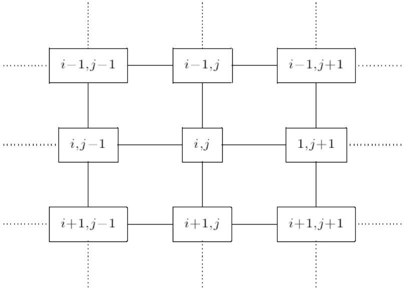
$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ -\sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) \right\},$$

  where Z, the normalizing constant, is called the partition function; $V_C(\mathbf{x}_C)$ are the clique potentials

- If $p(\mathbf{x})$ can be written in Gibbs form for the cliques of some graph, then it has the global Markov property.

Fundamental consequence: every Markov random field can be specified via clique potentials.

Regular rectangular lattices



- $\mathcal{V} = \{(i,j), i = 1,\ldots,M, j = 1,\ldots,N\}$
- Order-K neighborhood system:

$$\mathcal{N}^K(i,j) = \{(m,n) : (i-m)^2 + (j-n)^2 \leq K\}$$

Auto-models

- Only pair-wise interactions
- In terms of clique potentials: $|C| > 2 \Rightarrow V_C(\cdot) = 0$
- Simplest possible neighborhood models

Gauss-Markov Random Fields (GMRF)

- Joint probability function (assuming zero mean):

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1}\mathbf{x}\right\}$$

- Quadratic form in the exponent:

$$\mathbf{x}^T \boldsymbol{\Sigma}^{-1}\mathbf{x} = \sum_i \sum_j x_i x_i \Sigma_{i,j}^{-1} \Rightarrow \text{ auto-model}$$

- The neighborhood system is determined by the potential matrix $\boldsymbol{\Sigma}^{-1}$

- Local conditionals are univariate Gaussian

Gauss-Markov Random Fields (GMRF)

- Joint probability function (assuming zero mean):

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right\}$$

- Quadratic form in the exponent:

$$\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} = \sum_i \sum_j x_i x_i \Sigma_{i,j}^{-1} \Rightarrow \text{ auto-model}$$

- The neighborhood system is determined by the potential matrix $\mathbf{\Sigma}^{-1}$

- Local conditionals are univariate Gaussian

Gauss-Markov Random Fields

- Specification via clique potentials:

$$V_C(\mathbf{x}_C) = \frac{1}{2}\left(\sum_{i \in C} \alpha_i^C x_i\right)^2 = \frac{1}{2}\left(\sum_{i \in \mathcal{V}} \alpha_i^C x_i\right)^2$$

as long as $i \notin C \Rightarrow \alpha_i^C = 0$

- The exponent of the GMRF density becomes:

$$\begin{aligned} -\sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) &= -\frac{1}{2}\sum_{C \in \mathcal{C}}\left(\sum_{i \in \mathcal{V}} \alpha_i^C x_i\right)^2 \\ &= \frac{1}{2}\sum_{i \in \mathcal{V}}\sum_{j \in \mathcal{V}}\left(\sum_{C \in \mathcal{C}} \alpha_i^C \alpha_j^C\right) x_i x_j = -\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1}\mathbf{x}. \end{aligned}$$

GMRF: Application to image processing

- Classical image "smoothing" prior
- Consider an image to be a rectangular lattice with first-order pixel neighborhoods
- Cliques: pairs of vertically or horizontally adjacent pixels
- Clique potentials: squares of first-order differences (approximation of continuous derivative)

$$V_{\{(i,j),(i,j-1)\}}(x_{i,j}, x_{i,j-1}) = \frac{1}{2}(x_{i,j} - x_{i,j-1})^2$$

- Resulting $\mathbf{\Sigma}^{-1}$: block-tridiagonal with tridiagonal blocks

Bayesian image restoration with GMRF prior

- Observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \qquad \mathbf{n} \sim N(0, \sigma^2 \mathbf{I})$$

- Smoothing GMRF prior: $p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1}\mathbf{x}\}$
- MAP estimate:

$$\hat{\mathbf{x}} = [\sigma^2 \mathbf{\Sigma}^{-1} + \mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{y}$$

Bayesian image restoration with GMRF prior

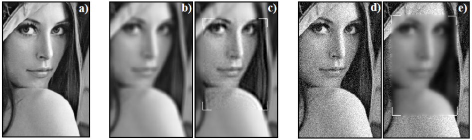


Figure: (a) Original image, (b) Blurred and slightly noisy image, (c) Restored version of (b), (d) No blur, severe noise, (e) Restored version of (d).

- Deblurring: good
- Denoising: oversmoothing; "edge discontinuities" smoothed out
- How to preserve discontinuities?
  - Other prior models
  - Hidden/latent binary random variables
  - Robust potential functions (e.g. L_2 vs. L_1-norm)

Bayesian image restoration with GMRF prior

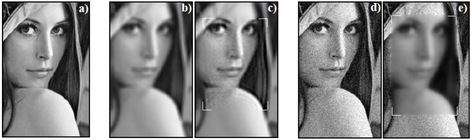


Figure: (a) Original image, (b) Blurred and slightly noisy image, (c) Restored version of (b), (d) No blur, severe noise, (e) Restored version of (d).

- Deblurring: good
- Denoising: oversmoothing; "edge discontinuities" smoothed out
- How to preserve discontinuities?
  - Other prior models
  - Hidden/latent binary random variables
  - Robust potential functions (e.g. L_2 vs. L_1-norm)

Compound GMRF

- Insert binary variables v to "turn off" clique potentials
- Modified clique potentials:

$$V(x_{i,j}, x_{i,j-1}, v_{i,j}) = \frac{1}{2}(1 - v_{i,j})(x_{i,j} - x_{i,j-1})^2$$

Intuitive explanation:

- $v = 0 \Rightarrow$ clique potential is quadratic ("on")
- $v = 1 \Rightarrow V_C(\cdot) = 0 \rightarrow$ no smoothing; image has an edge at this location.
- Can choose separate latent variables v and h for vertical and horizontal edges respectively

$$p(\mathbf{x}|\mathbf{h}, \mathbf{v}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v})\mathbf{x}\right\}$$

- MAP estimate:

$$\hat{\mathbf{x}} = [\sigma^2 \mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v}) + \mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{y}$$

Compound GMRF

- Insert binary variables v to "turn off" clique potentials
- Modified clique potentials:

$$V(x_{i,j}, x_{i,j-1}, v_{i,j}) = \frac{1}{2}(1 - v_{i,j})(x_{i,j} - x_{i,j-1})^2$$

Intuitive explanation:

- $v = 0 \Rightarrow$ clique potential is quadratic ("on")
- $v = 1 \Rightarrow V_C(\cdot) = 0 \rightarrow$ no smoothing; image has an edge at this location.
- Can choose separate latent variables v and h for vertical and horizontal edges respectively

$$p(\mathbf{x}|\mathbf{h}, \mathbf{v}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v})\mathbf{x}\right\}$$

- MAP estimate: $\hat{\mathbf{x}} = [\sigma^2\mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v}) + \mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{y}$

Compound GMRF

- Insert binary variables v to "turn off" clique potentials
- Modified clique potentials:

$$V(x_{i,j}, x_{i,j-1}, v_{i,j}) = \frac{1}{2}(1 - v_{i,j})(x_{i,j} - x_{i,j-1})^2$$

Intuitive explanation:

- $v = 0 \Rightarrow$ clique potential is quadratic ("on")
- $v = 1 \Rightarrow V_C(\cdot) = 0 \rightarrow$ no smoothing; image has an edge at this location.
- Can choose separate latent variables v and h for vertical and horizontal edges respectively

$$p(\mathbf{x}|\mathbf{h}, \mathbf{v}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v})\mathbf{x}\right\}$$

- MAP estimate:

$$\hat{\mathbf{x}} = [\sigma^2\mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v}) + \mathbf{H}^T\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{y}$$

Compound GMRF

- Insert binary variables v to "turn off" clique potentials
- Modified clique potentials:

$$V(x_{i,j}, x_{i,j-1}, v_{i,j}) = \frac{1}{2}(1 - v_{i,j})(x_{i,j} - x_{i,j-1})^2$$

Intuitive explanation:

- $v = 0 \Rightarrow$ clique potential is quadratic ("on")
- $v = 1 \Rightarrow V_C(\cdot) = 0 \rightarrow$ no smoothing; image has an edge at this location.
- Can choose separate latent variables v and h for vertical and horizontal edges respectively

$$p(\mathbf{x}|\mathbf{h}, \mathbf{v}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v})\mathbf{x}\right\}$$

- MAP estimate: $\hat{\mathbf{x}} = [\sigma^2 \mathbf{\Sigma}^{-1}(\mathbf{h}, \mathbf{v}) + \mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{y}$

Discontinuity-preserving restoration

Convex potentials:

- Generalized Gaussians: $V(x) = |x|^p, p \in [1, 2]$
- Stevenson:

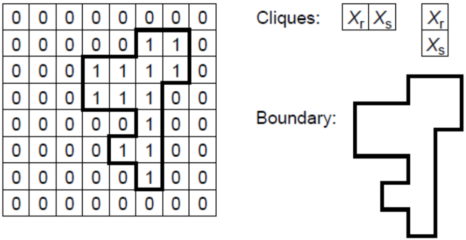
$$V(x) = \begin{cases} x^2, & |x| < a \\ 2a|x| - a^2, & |x| \geq a \end{cases}$$

- Green: $V(x) = 2a^2 \log\cosh(x/a)$

Non-convex potentials:

- Blake, Zisserman: $V(x) = (\min\{|x|, a\})^2$
- Geman, McClure: $V(x) = \frac{x^2}{x^2+a^2}$

Ising model (2-D MRF)

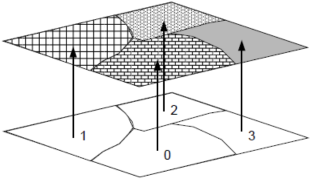


- $V(x_i, x_j) = \beta\delta(x_i \neq x_j), \beta$ is a model parameter
- Energy function:

$$\sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) = \beta(\text{boundary length})$$

- Longer boundaries less probable

Application: Image segmentation



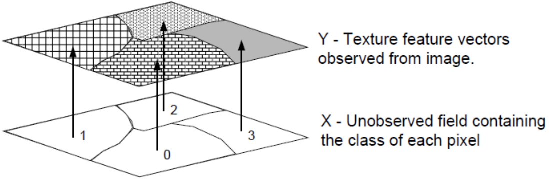
- Discrete MRF used to model the segmentation field
- Each class represented by a value $X_s \in \{0, \ldots, M-1\}$
- Joint distribution function:

$$P\{Y \in dy, X = x\} = p(y|x)p(x)$$

- (Bayesian) MAP estimation:

$$\hat{\mathrm{X}} = \arg\max_x p_{x|Y}(x|Y) = \arg\max_x (\log p(Y|x) + \log(p(x)))$$

Application: Image segmentation



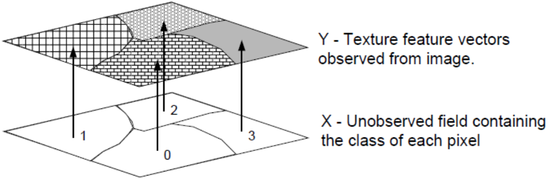
- Discrete MRF used to model the segmentation field
- Each class represented by a value $X_s \in \{0, \ldots, M-1\}$
- Joint distribution function:

$$P\{Y \in dy, X = x\} = p(y|x)p(x)$$

- (Bayesian) MAP estimation:

$$\hat{\mathbf{X}} = \arg\max_x p_{x|Y}(x|Y) = \arg\max_x(\log p(Y|x) + \log(p(x)))$$

Application: Image segmentation



- Discrete MRF used to model the segmentation field
- Each class represented by a value $X_s \in \{0, \ldots, M-1\}$
- Joint distribution function:

$$P\{Y \in dy, X = x\} = p(y|x)p(x)$$

- (Bayesian) MAP estimation:

$$\hat{\mathbf{X}} = \arg\max_x p_{x|Y}(x|Y) = \arg\max_x (\log p(Y|x) + \log(p(x)))$$

MAP optimization for segmentation

- Data model:

$$p_{y|x}(y|x) = \prod_{s \in S} p(y_s|x_s)$$

- Prior (Ising) model:

$$p_x(x) = \frac{1}{Z} \exp\{-\beta t_1(x)\},$$

where $t_1(x)$ is the number of horizontal and vertical neighbors of x having a different value

- MAP estimate:

$$\hat{x} = \arg\min_x \{-\log p_{y|x}(y|x) + \beta t_1(x)\}$$

- Hard optimization problem

MAP optimization for segmentation

- Data model:

$$p_{y|x}(y|x) = \prod_{s \in S} p(y_s|x_s)$$

- Prior (Ising) model:

$$p_x(x) = \frac{1}{Z} \exp\{-\beta t_1(x)\},$$

where $t_1(x)$ is the number of horizontal and vertical neighbors of x having a different value

- MAP estimate:

$$\hat{x} = \arg\min_x \{-\log p_{y|x}(y|x) + \beta t_1(x)\}$$

- Hard optimization problem

Some proposed approaches

- Iterated conditional modes: iterative minimization w.r.t. each pixel
- Simulated annealing: Generate samples from prior distribution
- Multi-scale resolution segmentation

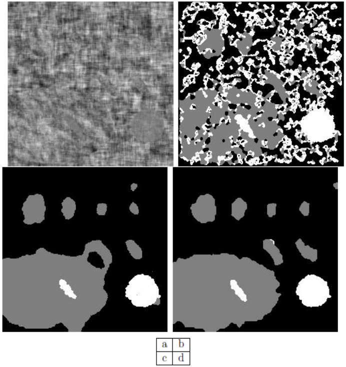


Figure: (a) Synthetic image with three textures, (b) ICM, (c) Simulated annealing, (d) Multi-resolution approach.

Summary

- Graphical models study probability distributions whose conditional dependencies arise out of specific graph structures
- Markov random field is an undirected graphical model with special factorization properties
- 2-D MRFs have been widely used as priors in image processing problems
- Choice of potential functions leads to different optimization problems