Markov Networks in Computer Vision

Sargur Srihari
srihari@cedar.buffalo.edu

Markov Networks for Computer Vision

- Some applications:
  1. Image segmentation
  2. Removal of blur/noise
  3. Stereo reconstruction
  4. Object recognition
- Typically called MRFs in vision community

1. Image Segmentation Task

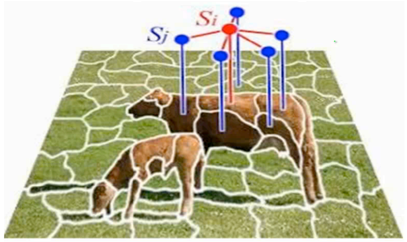
- Partition the image pixels into regions corresponding to distinct parts of scene
- Different variants of segmentation task
  - Many formulated as a Markov network
- Multiclass segmentation
  - Each variable X_i has a domain $\{1,..,K\}$ pixels
  - Value of X_i represents region assignment for pixel i, e.g., grass, water, sky, car
  - Classifying each pixel is expensive
    - Oversegment image into superpixels (coherent regions) and classify each superpixels
      - All pixels within region are assigned same value

Variables in Computer Vision

- X_i: Pixels or Super-pixels
- Joint probability distribution over an image
- Log-linear model:

$$\boxed{P(X_1,..X_n;\theta)=\frac{1}{Z(\theta)}\exp\left\{\sum_{i=1}^{k}\theta_i f_i(D_i)\right\}}$$

$$\boxed{\ln Z(\theta)=\ln\sum_{\xi}\exp\left\{\sum_i \theta_i f_i(\xi)\right\}}$$

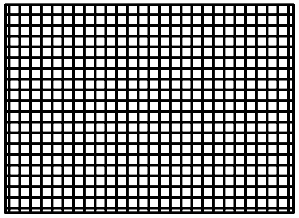


$\mathcal{F}=\{f_i\}$: Features between variables A,B ε D_i:

$$f_{a^0b^0}(a,b)=I\left\{a=a^0\right\}I\left\{b=b^0\right\}$$

Network Structure

- In most applications structure is pairwise
  - Variables correspond to pixels
  - Edges (factors) correspond to
    - interactions between adjacent pixels in grid on image
      - Each interior pixel has exactly four neighbors
    - Value space of variables and exact form of factors depend on task



- Usually formulated:
    - Factors in terms of energies
      - Negative log potentials
      - Values represent penalties:
        - » lower value = higher probability

Three Examples from Computer Vision

1. Image Segmentation

   Partition image pixels into regions

2. Image Denoising

   Restore "true" value of all pixels

3. Stereo Reconstruction

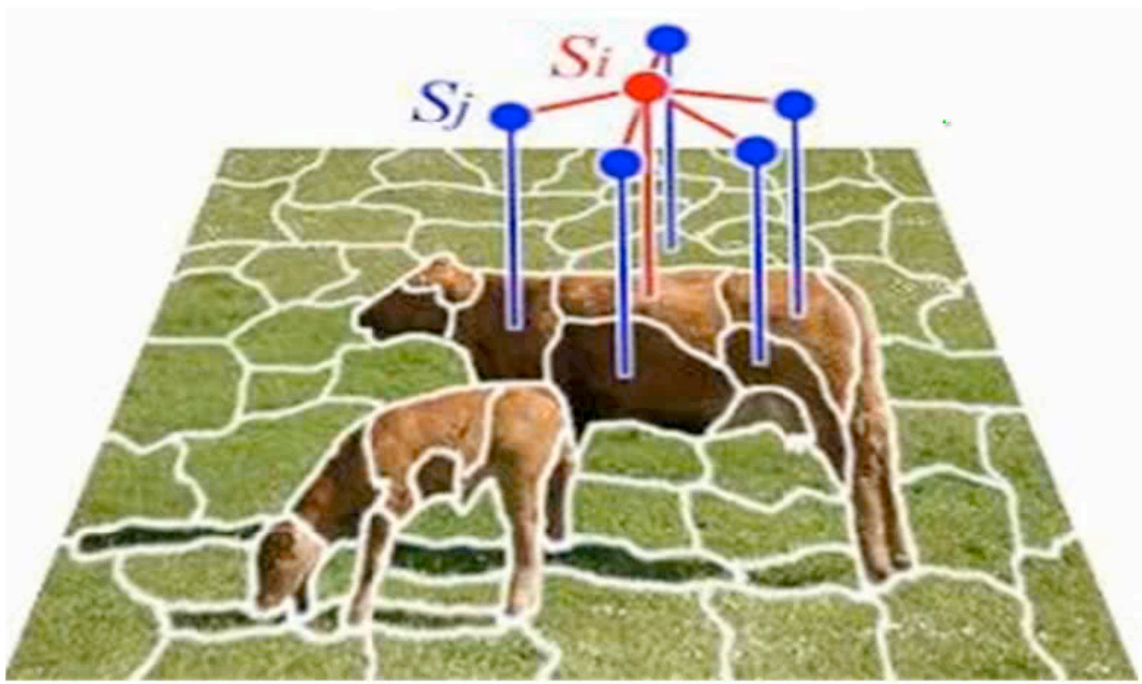
   Reconstruct depth disparity of each pixel

Model

- Edge potential between every pair of superpixels X_i, X_j
  - Encodes a contiguity preference
  - With a penalty λ whenever $X_i \neq X_j$
  - Model can be im[roved by making penalty depend on presence of an image gradient between pixels
  - Even better model:
    - Non default values for class pairs
    - Tigers adjacent to vegetation, water below vegetation

Graph from Superpixels

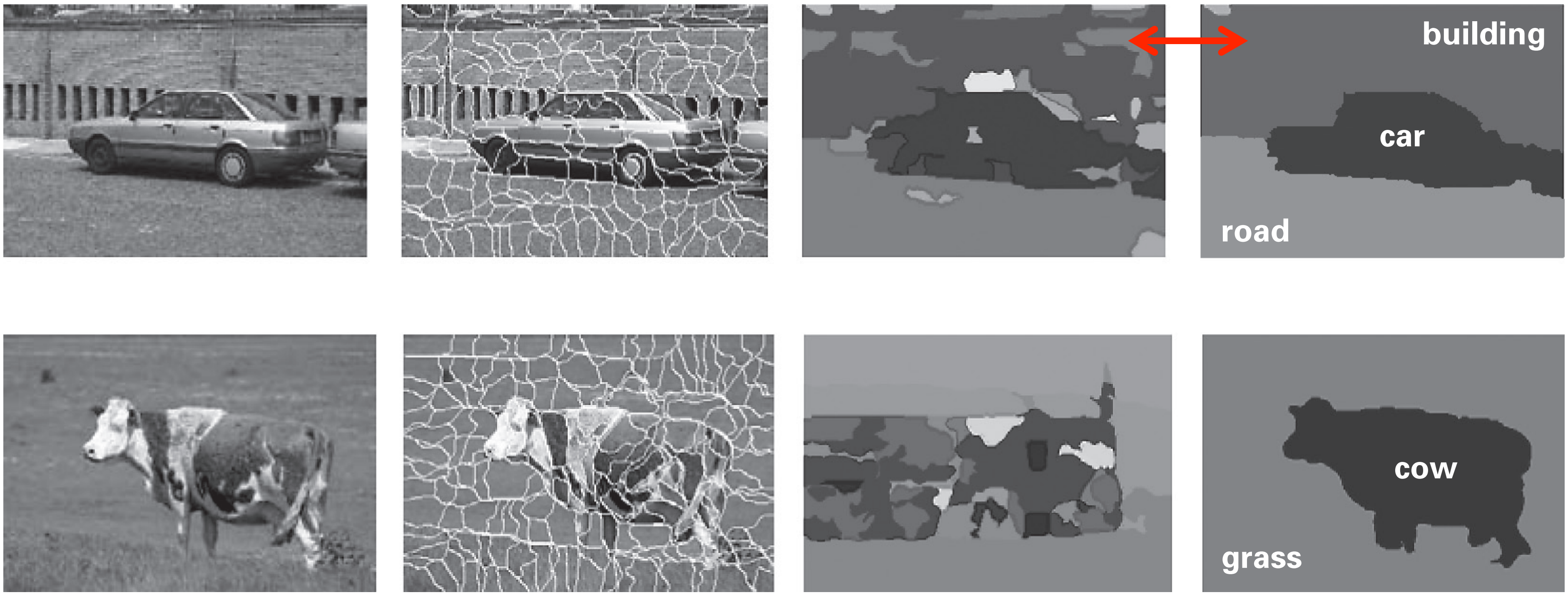
- A node for each superpixel
- Edge between nodes if regions are adjacent
- This defines a distribution in terms of this graph



Features for Image Segmentation

- Features extracted for each superpixel
  - Statistics over color, texture, location
    - Features either clustered or input to local classifiers to reduce dimensionality
    - Node potential is a function of these features
  - Factors depend upon pixels in image
  - Each image defines a different probability distribution over segment labels for pixels or superpixels
- Model in effect is a Conditional Random Field

Importance of Modeling Correlations between superpixels



Original image

Oversegmented image-superpixels Each superpixel is a random variable

Classification using node potentials alone-each superpixel classified independently

Segmentation using pairwise Markov Network encoding interactions between adjacent superpixels

Metric MRFs

- Class of MRFs used for labeling
- Graph of nodes $X_1,..X_n$ related by set of edges E
- Wish to assign to each X_i a label in space $V=\{v_1,..v_k\}$
- Each node, taken in isolation, has its preference among possible labels
- Also need to impose a soft"smoothness" constraint that neighboring nodes should take similar values

Encoding preferences

- Node preferences are node potentials in pairwise MRF
- Smoothness preferences are edge potentials
- Traditional to encode these models in negative log-space– using energy functions
- With MAP objective we can ignore the partition function

Energy Function

- Energy function

$$E(x_1,..x_n) = \sum_i \varepsilon_i(x_i) + \sum_{\{i,j\}} \varepsilon_{ij}(x_i x_j)$$

- Goal is to minimize the energy

$$\arg \min_{x_1,..x_n} E(x_1,..x_n)$$

Smoothness definition

- Slight variant of Ising model

$$\varepsilon_{ij}(x_1, x_j) = \begin{cases} 0 & x_i = x_j \\ \lambda_{i,j} & x_i \neq x_j \end{cases}$$

- Obtain lowest possible pairwise energy (0) when neighboring nodes X_i, X_j take the same value and a higher energy $\lambda_{i,j}$ when they do not

Generalizations

- Potts model extends it to more than two labels
- Distance function on labels
  - Prefer neighboring nodes to have labels smaller distance apart

Metric definition

- A function μ: $V \times V \rightarrow [0,\infty)$ is a metric if it satisfies
  - Reflexivity: $\mu(v_k,v_l)=0$ if and only if $k=l$
  - Symmetry: $\mu(v_k,v_l)=\mu(v_l,v_k)$;
  - Triangle Inequality: $\mu(v_k,v_l)+\mu(v_l,v_m) \geq \mu(v_k,v_m)$
- μ is a semi-metric if it satisfies first two
- Metric MRF is defined by defining

$$\varepsilon_{i,j}(v_k,v_j) = \mu(v_k,v_l)$$

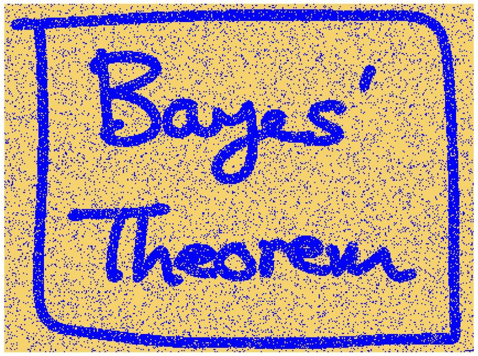
- A common metric: $\varepsilon(x_i,x_j) = \min(c||x_i\text{-}x_j||, dist_{max})$

2. Image denoising

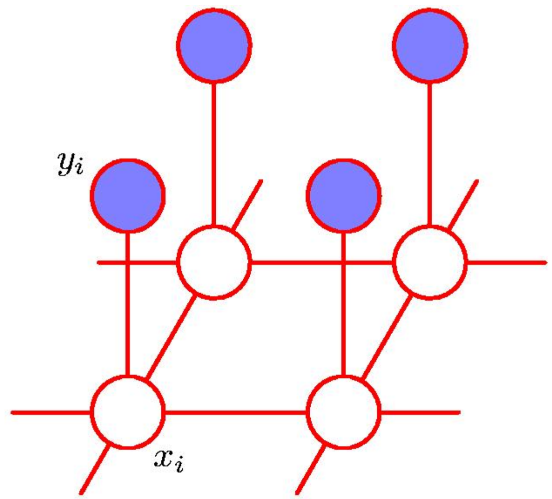
- Task: Restore true value given noisy pixel values
- Node potential $\phi(X_i)$ for each pixel X_i
    - penalize large discrepancies from observed pixel value y_i
- Edge potential
    - Encode continuity between adjacent pixel values
        - Penalize cases where inferred value of X_i is too far from inferred value of neighbor X_j
        - Important not to over-penalize true edge disparities (edges between objects or regions)
            - Leads to oversmoothing of image
        - Solution: Bound the penalty using a truncated norm
            - $\varepsilon(x_i,x_j) = \min(c\|x_i\text{-}x_j\|_{p,}\mathrm{dist}_{max})$ for $p \in \{1,2\}$

Binary Image de-noising

- Noise removal from binary image
- Observed noisy image
  - Binary pixel values $y_i \in \{-1,+1\}$, $i=1,..,D$
- Unknown noise-free image
  - Binary pixel values $x_i \in \{-1,+1\}$, $i=1,..,D$
- Noisy image assumed to randomly flip sign of pixels with small probability



Markov Random Field Model

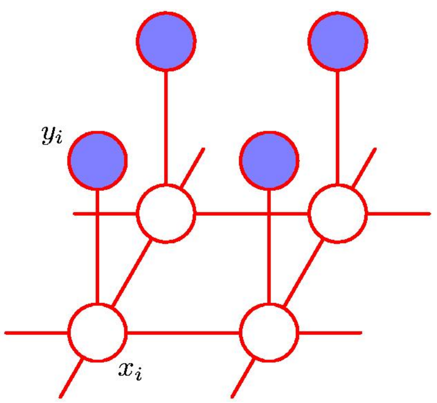


x_i= unknown noise-free pixel
y_i= known noisy pixel

- Known
  - Strong correlation between x_i and y_i
  - Neighbor pixels x_i and x_j are strongly correlated
- Prior knowledge captured using MRF

Energy Functions

- Graph has two types of cliques
  - $\{x_i, y_i\}$ expresses correlation between variables
    - Choose simple energy function $-\eta\, x_i y_i$
    - Lower energy (higher probability) when x_i and y_i have same sign
  - $\{x_i, x_j\}$ which are neighboring pixels
    - Choose $\beta x_i x_j$



Potential Function

- Complete energy function of model

$$E(\mathrm{x},\mathrm{y}) = h\sum_i x_i - \beta \sum_{\{i,j\}} x_i x_j - \eta \sum_i x_i y_i$$

  - The hx_i term biases towards pixel values that have one particular sign
  - Which defines a joint distribution over x and y given by

$$p(\mathrm{x},\mathrm{y}) = \frac{1}{Z}\exp\{-E(\mathrm{x},\mathrm{y})\}$$

De-noising problem statement

- We fix y to observed pixels in the noisy image
- $p(\mathrm{x}|\mathrm{y})$ is a conditional distribution over *all* noise-free images
  - Called *Ising* model in statistical physics
- We wish to find an image x that has a high probability

De-noising algorithm

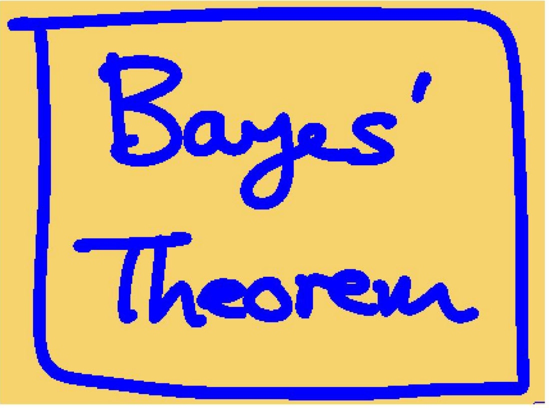
- Gradient ascent
  - Set x_i=y_i for all i
  - Take one node x_j at a time
    - evaluate total energy for states x_i=+1 and x_i=-1
    - keeping all other node variable fixed
  - Set x_j to value for which energy is lower
    - This is a local computation
    - which can be done efficiently
  - Repeat for all pixels until
    - a stopping criterion is met
  - Nodes updated systematically
    - by raster scan or randomly
- Finds a local maximum (which need not be global)
- Algorithm is called *Iterated Conditional Modes* (ICM)

Image Restoration Results

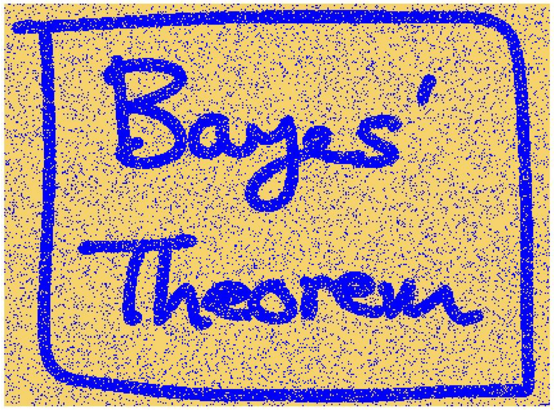
- Parameters

β=1.0, η=2.1, h=0

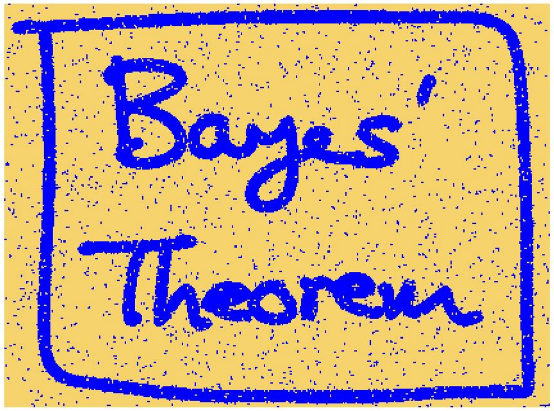
Noise Free image



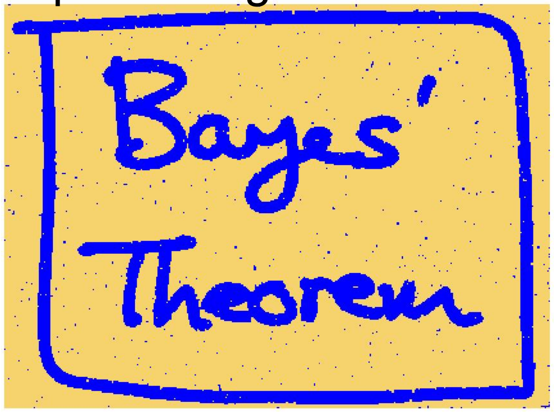
Noisy image where 10% of pixels are corrupted



Result of ICM



Global maximum obtained by Graph Cut algorithm



3. Stereo Reconstruction

- Reconstruct depth disparity of each pixel in the image
- Variables represent discretized version of depth dimension (more finely for discretized for close to camera and coarse when away)
- Node potential: a computer vision technique to estimate depth disparity
- Edge potential: a truncated metric
  - Inversely proportional to image gradient between pixels
    - Smaller penalty to large gradient suggesting occlusion