## Why Markov Random Field?

- Tasks on hand
  - Contextual constraints necessary in interpretation of visual information
  - Scene understanding in spatial and visual context of objects in it
  - Objects are recognized in the context of object features
- provides a convenient and consistent way of modeling context dependent entities such as image pixels and other spatially correlated features
- tells - how to model a priori probability of contextual dependent patterns such as a class of textures and an arrangement of object
- Visual modeling - Computational view point
- Define objective function for the optimal solution to a vision problem features



## Why MRF based Image Modeling?

- systematic and flexible treatment of the contextual information
- prior knowledge about the image labeling
  - image labellings possess a smoothness property of having contextual smoothness of the class labels in the image space so that a pixel with a particular class label is likely to share that label with its immediate neighbors
- used with statistical decision and estimation theories to formulate objective function in terms of optimality principles e.g. MAP-MRF
- Bayesian framework using MRF provides optimal solution
- optimization process using spatial local interactions makes parallel and local computations possible



## Markov Random Field: Applications

- MRF models for low level processing
  - Image restoration and segmentation
  - Surface reconstruction, Edge detection
  - Texture analysis, Shape from X, Optical flow
  - Data fusion, Active contours
- High level processing
  - Object matching, recognition
- MRF provides foundation for multi-resolution computation
- Equivalence between MRF and Gibbs distributions by Hammersley and Clifford (1971) and Besag (1974)
- Model mathematically sound and tractable in Bayesian framework Geman and Geman (1984)



## Applications of MRF Image Modeling

- Segmentation
  - used in image processing and computer vision
  - partition the image space into *distinctive* and *meaningful* homogeneous regions
  - random field $X$ in the MRF modeling is a set of region labels; taking one of $L_X$ integer values
  - realization of the observable random filed $Y$; a set of image graylevels
  - given image data $y$ find an optimal region label configuration $x^* = x^*(y)$ which is a best partition of $y$ with respect to some criterion
- medical, textured, document images - different optimization criteria

## Applications of MRF Image Modeling

- Noisy image segmentation
- regions with nearly constant gray levels assumed to be corrupted by iid Gaussian noise $Y_s = X_s + W_s \; \forall s \in \Omega$
- $W_s$ noise random variable with a zero mean white Gaussian distribution
- can be viewed as restoration problem; removing the white noise component $W = w$ in the observed image data $y$
- texture classification and segmentation
- each texture characterized by regular patterns
- observable image data $y$ in each texture spatially correlated - spatial treatment required for modeling the random field $Y$
- document image segmentation: segment into text, picture, graph and background regions



## Applications of MRF Image Modeling

- color image segmentation
- color difference of neighboring pixels is modeled by MRF
- medical image segmentation - different tissue classes
- biological image segmentation - biological cell images into three parts: original sample (center of image), nutritive substance (background of image) and the new cells
- synthetic aperture radar (SAR) image segmentation: speckle noise in SAR images, shadow like noise, classify into ice and water
- aerial image segmentation: classify into regions man-made and natural areas
- sonar image segmentation: speckle noise in sonar images
- low depth of field image segmentation: distinguish focused object from defocused background



## Applications of MRF Image Modeling

- image restoration (image denoising)
- recover original image data $x$ of the unobservable random field $X$ from its blurred and noisy observation $y$ of the observable random field $Y$
- to solve ill-posed problem - restrict the solution space by employing a smoothness constraint, edge-preserving modeling, in transform domain wavelet coefficient assumed to be corrupted with noise
- texture image synthesis by controlling the parameters (clique potential) associated with Gibbs distribution, or using Gaussian Markov random field
- estimating visual motion in video using smoothness motion constraint
- image retrieval
- face detection and recognition



## MRF: A Labeling Problem

- Vision problem posed as labeling problem
- Solution to a problem is a set to labels assigned to image pixels or features
- Labeling is a natural representation for the study of MRF
- Labeling problem is specified in terms of a set of sites and a set of labels
- $S$ is a discrete set of $m$ sites $S = \{1, \ldots, m\}$
- Site represents a point or a region in Euclidean space
  - an image pixel or an image feature such as a corner point
  - a line segment or a surface patch
- Sites categorized in terms of their regularity
- Sites not spatial regular considered as irregular e.g. features extracted from images

# MRF: Sites and Label

- Sites on a lattice are spatially regular
- Rectangular lattice for a 2D image of size $n \times n$

$$S = \{(i,j) | 1 \le i,j \le n\}$$

  $m$ sites where $m = n \times n$
- Inter-relationship between sites - neighborhood system
- A label is an event that may happen to a site $\mathcal{L}$ a set of labels
- Label set may continuous or discrete
- Discrete case $\mathcal{L} = \{1, \ldots, M\} = \{l_1, \ldots, l_M\}$
- Continuous label correspond to a real line $\mathcal{R}$ or an interval label may take a vector or matrix value



# MRF: Labeling

- It may be ordered set ($\{1, \ldots, 255\}$) or unordered set (texture types)
- for ordered label set a numerical (quantitative) measure of similarity between any two labels
- for unordered label a similarity measure is symbolic (qualitative) say equal or non-equal
- Labeling: What is it?
- Assign a label from $\mathcal{L}$ to each of the sites in $S$
- $f = \{f_1, \ldots, f_m\}$ called a labeling of sites in $S$ in terms of labels in $\mathcal{L}$
- $f_i = f(i)$ a function with domain $S$ and image $\mathcal{L}$ mapping from $S$ to $\mathcal{L}$

$$f : S \to \mathcal{L}$$



# MRF: Field and Configuration

- Labeling is also called a coloring in mathematical programming
- In terminology of random fields, a labeling is called a configuration
- In vision, a configuration or labeling can correspond to an image, an edge map, an interpretation of image features in terms of object features
- All sites have the same label set, the set of all possible labellings, i.e, configuration space is Cartesian product

$$\mathcal{F} = \mathcal{L} \times \cdots \times \mathcal{L} = \mathcal{L}^m$$

- For example, $\mathcal{L}$ is possible pixel value set then $\mathcal{F}$ defines all admissible images
- In general every site $i$ may have its own admissible set $\mathcal{L}_i$

$$\mathcal{F} = \mathcal{L}_1 \times \mathcal{L}_2 \cdots \times \mathcal{L}_m$$



# MRF: Sites and Label Category

- LP1: Regular sites - continuous label
- LP2: Regular sites - discrete label
- LP3: Irregular sites - discrete label
- LP4: Irregular sites - continuous label
- First two for low level processing on images
- Other two for high level processing on features
- Restoration or smoothing of images - LP1 or LP2 depending upon pixel values continuous or discrete
- Region segmentation, Edge detection - LP2
- Perceptual grouping of detected features - grouped into connected or disconnected - LP3
- Feature based object matching and recognition - LP3
- Pose estimation - LP4

## MRF: Contextual Constraints

- Contextual constraints expressed locally in terms of conditional probabilities $P(f_i|\{f_{i'}\})$
- $\{f_{i'}\}$ denotes the set of labels at the other sites $i' \neq i$
- or globally as the joint probability $P(f)$
- If labels are independent one another (no context)
  - the joint probability is the product of the local ones

$$P(f) = \prod_{i \in S} P(f_i)$$

  - conditional independence

$$P(f_i|\{f_{i'}\}) = P(f_i) \quad i' \neq i$$

  - A global labeling $f$ can be computed by considering each label $f_i$ locally.



## MRF: Neighborhood System

- If labels are mutually dependent then how to make a global inference using local information becomes a non-trivial task
- MRF provides a mathematical foundation for solving this problem.
- MRF and Gibbs Distribution: a theory to analyze the spatial or contextual dependencies of physical phenomena
- Based on Neighborhood System and Cliques
- A neighborhood system is defined as

$$\mathcal{N} = \{\mathcal{N}_i | \forall i \in S\}$$

  $\mathcal{N}_i$ the set of sites neighboring $i$
- The neighboring relationship has the following properties
  - a site is not neighboring to itself $i \notin \mathcal{N}_i$
  - the neighboring relationship is mutual $i \in \mathcal{N}_{i'} \Rightarrow i' \in \mathcal{N}_i$



## MRF: Types of Neighborhood System

- For a regular lattice $S$, the neighbor set of $i$ is defined as the set of nearby sites within a radius $r$

$$\mathcal{N}_i = \{i' \in S | [dist(pixel_{i'}, pixel_i)]^2 \leq r, i' \neq i\}$$

- First order neighborhood system called 4-neighborhood system, every (interior) site has four neighbors
- Second order neighborhood system called 8-neighborhood system
- $n^{th}$ order neighborhood system
- Sites in a regular rectangular lattice $S = \{(i,j) | 1 \leq i,j \leq n\}$ four nearest neighbor $N_{i,j} = \{(i-1,j),(i+1,j)(i,j-1),(i,j+1)\}$
- For an irregular $S$

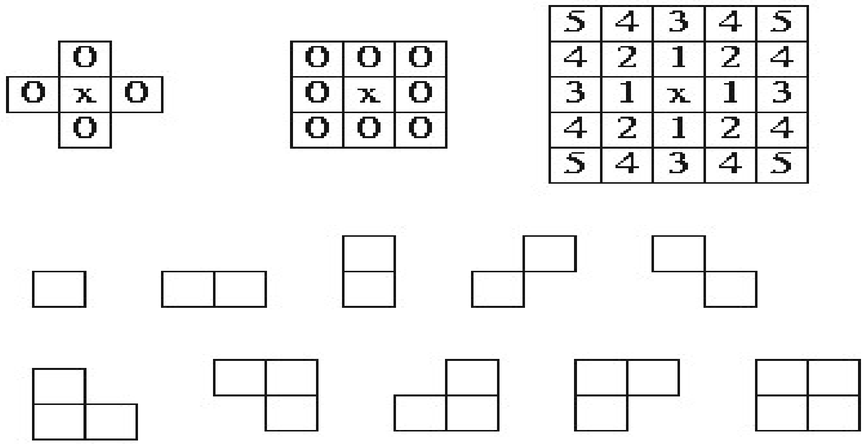
$$\mathcal{N}_i = \{i' \in S | [dist(feature_{i'}, feature_i)]^2 \leq r, i' \neq i\}$$



## MRF: Cliques

- In general, the neighbor set $\mathcal{N}_i$ for an irregular $S$ have varying shapes and sizes
- A clique $c$ for $(S, \mathcal{N})$ is defined as a subset of sites in $S$
- single site $\mathcal{C}_1 = \{i | i \in S\}$
- pair of neighboring sites $\mathcal{C}_2 = \{\{i, i'\} | i' \in \mathcal{N}_i, i \in S\}$
- triple of neighboring sites $\mathcal{C}_3 = \{\{i, i', i"\} | i, i', i" \in S\}$ are neighbors to one another
- Sites in a clique are ordered and $\{i, i'\}$ is not the same clique as $\{i', i\}$
- The collection of all cliques for $(S, \mathcal{N})$ $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \ldots$
- As the order of the neighborhood system increases, the number of cliques grow rapidly and computationally expensive
- Cliques for irregular sites do not have fixed shapes as those for a regular lattice and their types are depicted by the number of involved sites

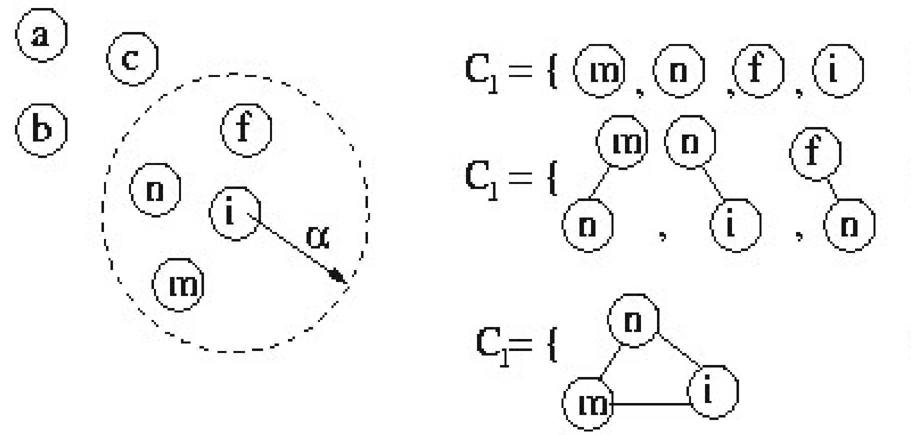
# MRF: Neighborhood and Cliques on Regular Site



- Upper: Neighborhood 4, 8, ... $n^{th}$
- Lower: Cliques



# MRF: Neighborhood and Cliques on Irregular Site




- Left: Neighborhood
- Right: Cliques



# MRF: A Random Field

- Let $F = \{F_1, \ldots, F_m\}$ be a family of random variables defined on the set $S$
- Each random variable $F_i$ takes a value $f_i$ in $\mathcal{L}$
- The family $F$ is called a random field
- $F_i = f_i$ denotes the event that $F_i$ takes the value $f_i$
- $(F_1 = f_1, \ldots, F_m = f_m)$ to denote the joint event
- A joint event is abbreviated as $F = f$ where $f = \{f_1, \ldots, f_m\}$ is a configuration of $F$, corresponding to a realization of the field
- For a discrete label set $\mathcal{L}$, the probability that random variable $F_i$ takes the value $f_i$ is denoted by $P(F_i = f_i)$, abbreviated $P(f_i)$
- The joint probability is denoted $P(F = f) = P(F_1 = f_1, \ldots, F_m = f_m)$ and abbreviated $P(f)$



# MRF: Requisite Conditions

- For a continuous $\mathcal{L}$, the same is represented by probability density functions, $p(F_i = f_i)$ and $p(F = f)$
- $F$ is said to be a Markov random field on $S$ with respect to a neighborhood system $\mathcal{N}$ if and only if the following two conditions are satisfied:
  1. $P(f) > 0, \forall f \in \mathbb{F}$ (positivity)
  2. $P(f_i | f_{S-\{i\}}) = P(f_i | f_{\mathcal{N}_i})$ (Markovianity)

  where $S - \{i\}$ is the set difference, $f_{S-\{i\}}$ denotes the set of labels at the sites in $S - \{i\}$ and

$$f_{\mathcal{N}_i} = \{f_{i'} | i' \in \mathcal{N}_i\}$$

  stands for the set of labels at the sites neighboring $i$

## MRF: Other Property

- The positivity condition is assumed because the joint probability $P(f)$ of any random field is uniquely determined by its local conditional probabilities, if positivity condition is satisfied
- The Markovianity depicts the local characteristics of $F$. The label at a site is dependent only on those at the neighboring sites, i.e, neighboring labels have direct interactions on each other
- MRF have other properties as homogeneity and isotropy. It is said to be homogeneous if $P(f_i|f_{N_i})$ is regardless of the relative position of site $i$ in $S$.
- Equivalence between Markov random fields and Gibbs distribution provides a mathematically tractable means of specifying the joint probability of an MRF.



## Gibbs Random Fields

- A set of random variables $F$ is said to be a Gibbs random field (GRF) on $S$ with respect to $\mathcal{N}$ if and only if its configurations obey a Gibbs distribution
- A Gibbs distribution takes the following form

$$P(f) = Z^{-1} \times e^{-\frac{U(f)}{T}} \qquad Z = \sum_{f \in \mathbb{F}} e^{-\frac{U(f)}{T}}$$

  is a normalizing constant called the partition function, $T$ is a constant called the temperature which shall be assumed to be 1 and $U(f)$ is the energy function.
- The energy

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \qquad (1)$$

  is a sum of clique potentials $V_c(f)$ over all possible cliques $\mathcal{C}$
- The value of $V_c(f)$ depends on the local configuration on the clique $c$



## Gibbs Random Fields: Property

- The Gaussian distribution is a special member of this Gibbs distribution family
- A GRF is said to be homogeneous if $V_c(f)$ is independent of the relative position of the clique $c$ in $S$
- It is said to be isotropic if $V_c$ is independent of the orientation of $c$
- To calculate a Gibbs distribution, it is necessary to evaluate the partition function $Z$ which is the sum over all possible configuration in $\mathbb{F}$
- There are a combinatorial number of elements in $\mathbb{F}$ for a discrete $\mathcal{L}$, the evaluation is prohibitive even for problems of moderate sizes
- Several approximation methods exist for solving this problem.



## Gibbs Random Fields: Configuration

- $P(f)$ measures the probability of the occurrence of a particular configuration or pattern, $f$
- The more probable configurations are those with lower energies
- The temperature $T$ controls the sharpness of the distribution
- When the temperature is high, all configurations tend to be equally distributed
- Near the zero temperature, the distribution concentrates around the global energy minima
- Given $T$ and $U(f)$, a class of patterns can be generated by sampling the configuration space $\mathbb{F}$ according to $P(f)$
- For discrete labeling problem, a clique potential $V_c(f)$ can be specified by a number of parameters
- For continuous labeling problem, $V_c(f)$ is continuous function of $f_c$

## Gibbs Random Fields: Energy

- The energy of a Gibbs distribution is expressed as the sum of several terms, each related to cliques of a certain size, i.e.,

$$U(f) = \sum_{\{i\}\in\mathcal{C}_1} V_1(f_i) + \sum_{\{i,i'\}\in\mathcal{C}_2} V_2(f_i, f_{i'}) + \sum_{\{i,i',i''\}\in\mathcal{C}_3} V_2(f_i, f_{i'}, f_{i''}) + \dots$$

- An important special case is when only cliques of size up to two are considered. In this case, the energy can also be written as

$$U(f) = \sum_{i\in S} V_1(f_i) + \sum_{i\in S}\sum_{i'\in\mathcal{N}_i} V_2(f_i, f_{i'})$$

- $\{i, i'\}$ and $\{i', i\}$ are two distinct cliques in $\mathcal{C}_2$ because the sites in a clique are ordered



## GRF: Equivalence

- The conditional probability can be written as

$$P(f_i|f_{\mathcal{N}_i}) = \frac{e^{-\left[V_1(f_i)+\sum_{i'\in\mathcal{N}_i} V_2(f_i,f_{i'})\right]}}{\sum_{f_i\in\mathcal{C}} e^{-\left[V_1(f_i)+\sum_{i'\in\mathcal{N}_i} V_2(f_i,f_{i'})\right]}}$$

- Equivalence between Markov Random Field and Gibbs Random Field
- An MRF is characterized by its local property (the Markovianity)
- A GRF is characterized by its global property (the Gibbs distribution)
- The Hammersley-Clifford theorem establishes the equivalence of these two types of properties
- The theorem states that $F$ is an MRF on $S$ with respect to $\mathcal{N}$ if and only if $F$ is a GRF on $S$ with respect to $\mathcal{N}$.



## Markov-Gibbs Equivalence

- This theorem provides a simple way of specifying the joint probability
- *One can specify the joint probability $P(F = f)$ by specifying the clique potential functions $V_c(f)$ and choose appropriate potential functions for desired system behavior. It encodes the a priori knowledge or preference about interactions between labels*
- *The MRF-Gibbs distribution has been used in solving optimization problems. In optimization problems, an objective function is in the form of an energy function and is to be minimized*
- *As the quantitative cost measure, an energy function defines the minimal solution as its minimum, usually a global one*
- *Formulate an energy function so that the correct solution is embedded as the minimum*



## MRF Image Modeling: Example

- image modeled as a random field $Y = \{Y_s : s \in L\}$
- $L = \{(i,j)|0 \le i \le N_1 - 1, 0 \le j \le N_2 - 1\}$ index set, a set of site indices on 2-D discrete $N_1 \times N_2$ rectangular integer lattice
- for each lattice point or pixel $s = (i,j) \in S$ $Y_s$ is a real-valued random variable
- random field $Y$ is characterized by a joint probability distribution $P_Y$ which may be characterized by an associated parameter set $\theta_Y$
- random variables $Y_s$ will take on sample values or realizations $y_s$ from a common finite set of integers $\{0, 1, 2, \dots, L_Y - 1\}$
- $y_s$ may be feature values depending on the application; DCT or DWT coefficients

## MRF Image Modeling: Example

- let random field $X = \{X_1, \ldots, X_m\}$ be a Markov Random Field defined on $L$ and $m$ is the total number of the classes
- sites in $L$ are related to each other via a neighborhood system $\Psi = \{N_l, l \in L\}$ $N_l$ is the set of neighbor of site
- site is not a neighbor of itself
- clique is a subset of sites in $N_l$ $c \in N_l$ is a clique of every pair of distinct sites in $c$ are neighbors
- random field $X$ is considered to be an MRF on $S$ if and only if $P(X = x) > 0$ and $P(X_l = x_l | X_r = x_r, r \neq s) = P(X_l = x_l | X_r = x_r, r \in N_l)$
- difficult to determine the above characteristics in practice



## MRF Image Modeling: Example

- MRF has the form

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

- $X = x$ is a realization from $X = \{X_1, \ldots, X_m\}$, i.e., $x = \{x_1, \ldots, x_m\}$ is a set of random field $X$ and

$$U(X) = \sum_L V_c(x)$$

  is global energy function and it is given by the sum of clique potentials $V_c(x)$, over all possible cliques
- choice of energy function is arbitrary
- for example, several definition of $U(X)$ in the framework of image segmentation



## MRF Image Modeling: Example

- general expression for the energy function

$$U(X) = \sum_L V_c(ci) + \sum_{c \in N_l} V_c(ci, cj)$$

- known as Potts model $V_c(ci)$ external field that weighs the relative importance of different classes
- simplified Potts model with no external energy $V_c(ci) = 0$
- local spatial transitions are taken into account and all the classes in the label segmentation, $X_{opt}$, as near as possible to the real image $X^*$

$$V_c(ci, cj) = \begin{cases} -\beta & \textit{if } ci = cj \\ 0 & \textit{otherwise} \end{cases}$$

$$V_c(ci, cj) = \begin{cases} -\frac{\beta \sigma_i^2}{(\sigma_i^2 + (y_{ci} - y_{cj})^2 \times d_{cicj})} & \textit{if} \quad ci \neq cj \\ -\beta & \textit{if} \quad ci = cj \end{cases}$$



## MRF-MAP Segmentation

- $y_{ci}$ and $y_{cj}$ are the pixel intensities of $ci$ and $cj$, $d_{cicj} = 1$ $or \sqrt{2}$ represents the distance between the two pixels
- $\beta$ is constant that controls the classification
- decreasing of $d_{cicj}$ and $(y_{ci} - y_{cj})^2$, $V_c(ci, cj)$ decreased to $-\beta$, i.e., $ci$ and $cj$ are right one pixel or their intensities are same
- image $Y$ as rectangular lattice $L$ $y_l$ denotes the intensity of the pixel at $l$ and it correspond to the label $x_l$ in $X$
- Bayes theorem yields a complete model coupling intensities and labels

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

- $P(X|Y)$ is the posteriori, $P(Y|X)$ is the conditional probability density of the image $Y$, $P(X)$ is the prior density of labelling $X$

## MRF-MAP Segmentation

- the prior probability of the image $P(Y)$ is independent of the labelling $X$, using MAP

$$X_{opt} = \max \arg_{X \in L}\{P(Y|X)P(X)\}$$

- assume that image data is obtained by adding an identical independently distributed (i.i.d) Gaussian noise

$$p(y_l|x_l) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left\{-\frac{(y_l - \mu_{xl})^2}{2\sigma_l^2}\right\}$$

- based on conditional independent assumption of $Y$ the conditional density $P(Y|X) = \prod_L P(y_l|x_l)$



## MRF-MAP Segmentation

$$P(Y|X) = \prod_L \left[\frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left\{-\frac{(y_l - \mu_{xl})^2}{2\sigma_l^2}\right\}\right] \propto \exp\left[-\sum_L \frac{(y_l - \mu_{xl})^2}{2\sigma_l^2}\right]$$

- potential function of the conditional probability can be written as

$$U(Y|X) = \sum_L \frac{(y_l - \mu_{xl})^2}{2\sigma_l^2}$$

- similarly prior density of MRF takes the form of

$$P(X) = \frac{1}{Z} \exp\left[-\sum_L \sum_{c \in N_l} \frac{V_c(ci, cj)}{T}\right]$$

$$P(X|Y) \propto \exp\left(-\sum_L \left[\frac{(y_l - \mu_{xl})^2}{2\sigma^2} + \sum_{c \in N_l} \frac{V_c(ci, cj)}{T}\right]\right)$$



## MRF-MAP Segmentation

- image $Y$ is then segmented by finding the field of labels $X$

$$\begin{aligned} X_{opt} &= \min \arg_{x \in L} U(X|Y) \\ &= \min \arg \left(\sum_L \left[\frac{(y_l - \mu_l)^2}{2\sigma^2} + \frac{1}{T}\sum_{c \in N_l} V_c(ci, cj)\right]\right) \end{aligned}$$

- optimization method: simulated annealing (SA), iterated conditional model (ICM) used to find the solution
- SA is slow but guarantee a global minimum solution
- ICM likely to reach local minima and no guarantee that a global minimum of energy function can be obtained, provides much faster convergence
- ICM iteratively decrease the energy by visiting and updating the pixel



## MRF-MAP Segmentation

- for each pixel $l$, given the observed image $Y$ and current labels of all the pixels in the neighborhood, the label of $X_l$ is replaced with one that can maximize the probability as

$$X_l^{(k+1)} = \arg\max P(X_l^{(k)}|Y, X_r^{(k)}, r = l)$$

- starting from the initial state, keep on running on the procedure above until either the predefined number of iterations is reached or the label of $X$ does not change