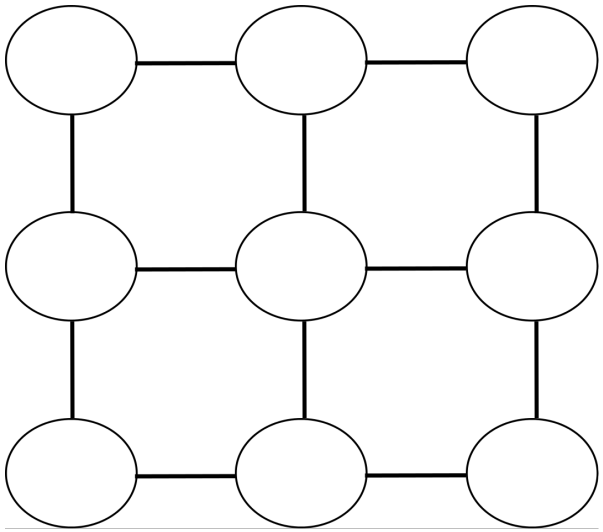
Markov Random Fields

Probabilistic Graphical Models

L. Enrique Sucar, INAOE



Outline

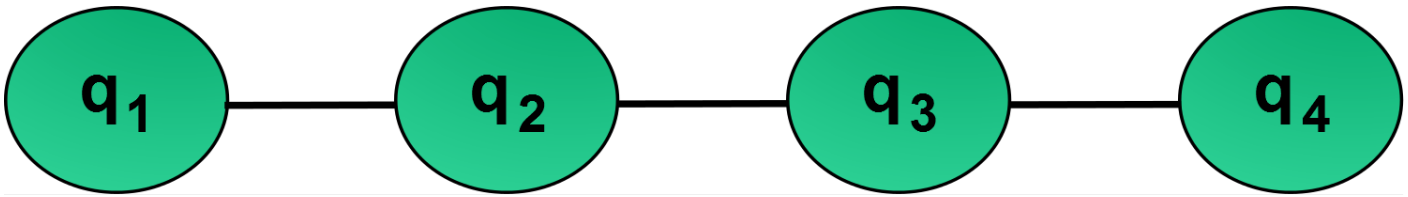
Introduction

- Certain processes, such as a ferromagnetic material under a magnetic field, or an image, can be modeled as a series of *states* in a chain or a regular grid
- Each state can take different values and is influenced probabilistically by the states of its neighbors
- These models are known as *Markov random fields* (MRFs)

Ising Model

- In an Ising model, there are a series of random variables in a line; each random variable represents a dipole that could be in two possible states, *up* (+) or *down* (-)
- The state of each dipole depends on an external field and the state of its neighbor dipoles in the line
- A *configuration* of a MRF is a particular assignment of values to each variable in the model
- A MRF is represented as an undirected graphical model

Example



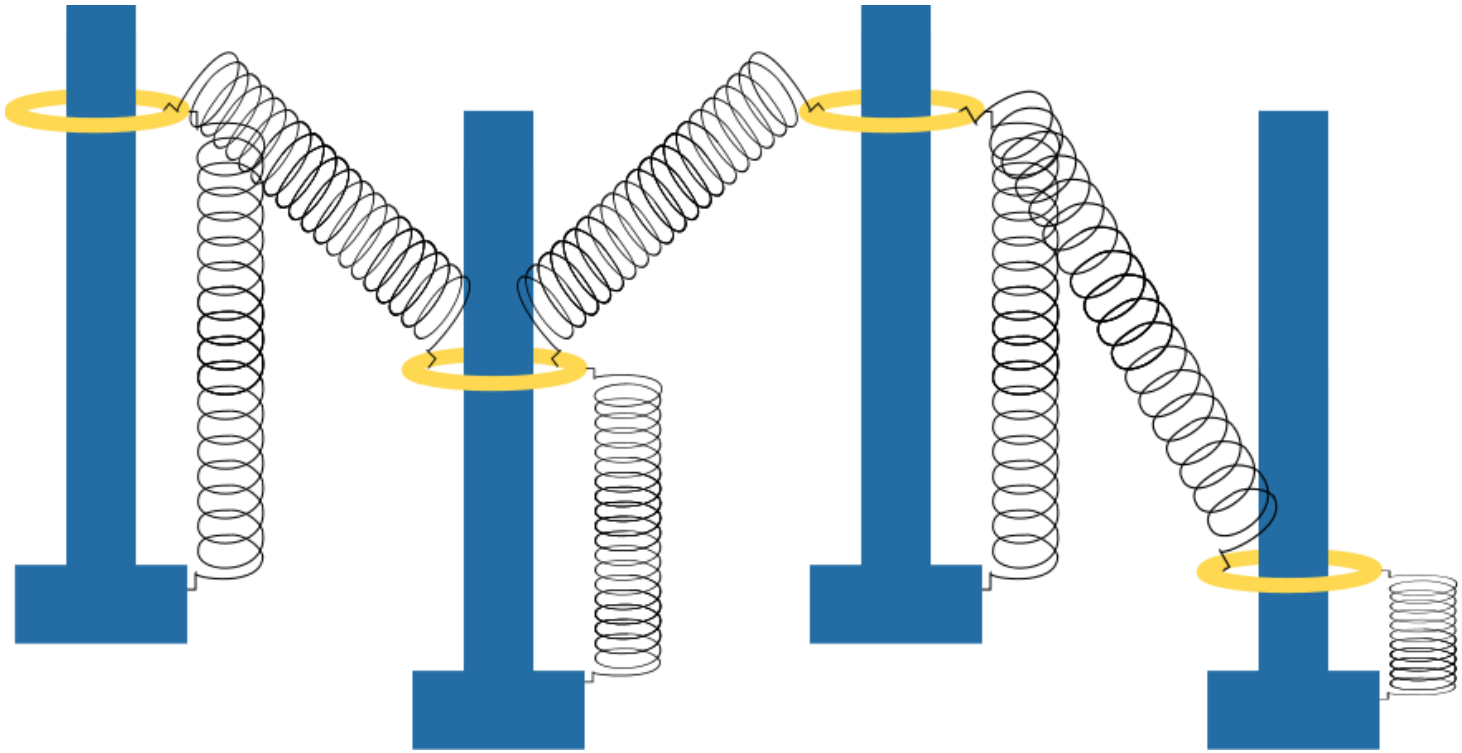
Properties and Central Problem

- An important property of a MRF is that the state of a variable is independent of all other variables in the model given its neighbors in the graph
- The central problem in a MRF is to find the configuration of maximum probability
- The probability of a configuration depends on the combination of an *external* influence (e.g., a magnetic field in the Ising model) and the *internal* influence of its neighbors

Physical Analogy

- A MRF can be thought of as a series of rings in poles, where each ring represents a random variable, and the height of a ring in a pole corresponds to its state
- Each ring is attached to its neighbors with a spring, this corresponds to the internal influences; and it is also attached to the base of its pole with another spring, representing the external influence
- The relation between the springs' constants defines the relative weight between the internal and external influences
- If the rings are left loose, they will stabilize to a configuration of minimum energy, that corresponds to the configuration with maximum probability

Physical Analogy



Random Fields

- A *random field* (RF) is a collection of S random variables, $\mathbf{F} = F_1, \ldots F_s$, indexed by sites
- Random variables can be discrete or continuous
- In a discrete RF, a random variable can take a value f_i from a set of m possible values or labels $L = \{l_1, l_2, \ldots l_m\}$

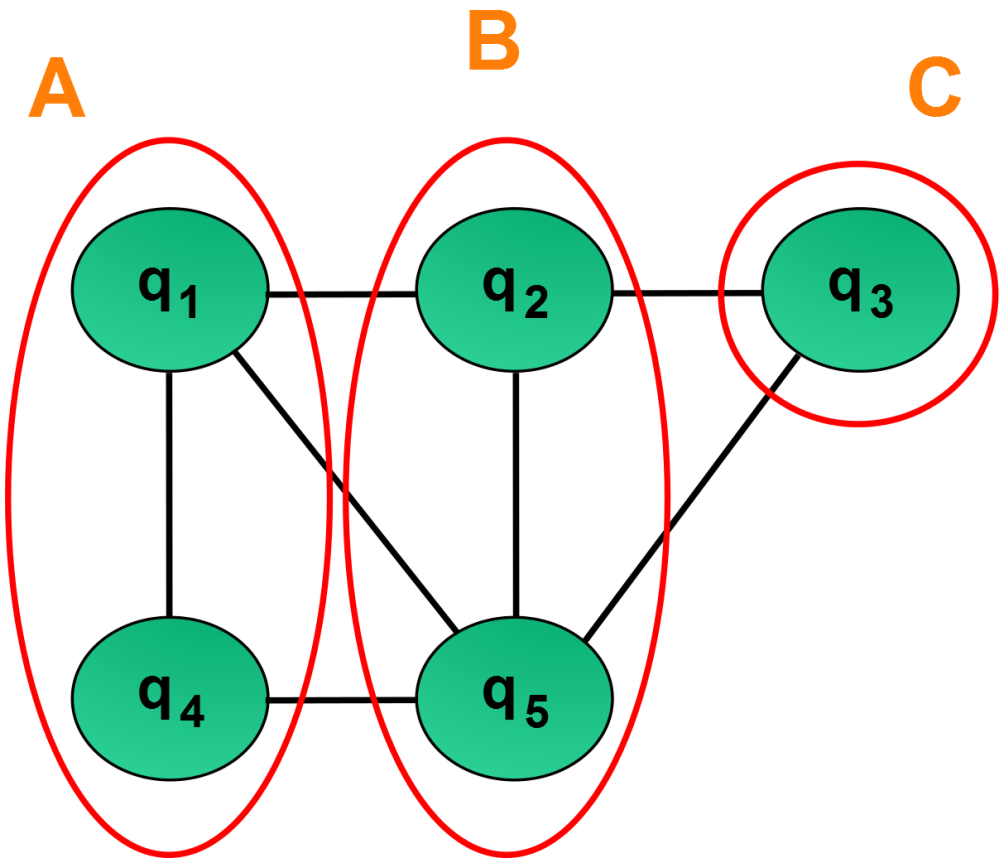
Markov Random Field

- A *Markov random field* or Markov network (MN) is a random field that satisfies the locality property: a variable F_i is independent of all other variables in the field given its neighbors:

$$P(F_i \mid \mathbf{F_c}) = P(F_i \mid Nei(F_i)) \tag{1}$$

- Graphically, a Markov network (MN) is an undirected graphical model which consists of a set of random variables, **V**, and a set of undirected edges, **E**

Independence Relations



- A subset of variables **A** is independent of the subset of variables **C** given **B**, if the variables in **B** *separate* **A** and **C** in the graph

Joint Probability Distribution

- The joint probability of a MN can be expressed as the product of local functions on subsets of variables
- These subsets should include, at least, all the *cliques* in the network
- For the example:
  $P(q_1, q_2, q_3, q_4, q_5) =$
  $(1/k)P(q_1, q_4, q_5)P(q_1, q_2, q_5)P(q_2, q_3, q_5)$

Definition

- A Markov network is a set of random variables, $\mathbf{X} = X_1, X_2, ..., X_n$ that are indexed by V, such that $G = (V, E)$ is an undirected graph, that satisfies the Markov property
- A variable X_i is independent of all other variables given its neighbors, $Nei(X_i)$:

$$P(X_i \mid X_1, ...X_{i-1}, X_{i+1}, ..., X_n) = P(X_i \mid Nei(X_i)) \quad (2)$$

- The neighbors of a variable are all the variables that are directly connected to it

Factorization

- Under certain conditions (if the probability distribution is strictly positive), the the joint probability distribution of a MRF can be factorized over the cliques of the graph:

$$P(\mathbf{X}) = (1/k) \prod_{C \in Cliques(G)} \phi_C(X_C) \tag{3}$$

- A MRF can be categorized as *regular* or *irregular*. When the random variables are in a lattice it is considered regular; for instance, they could represent the pixels in an image; if not, they are irregular.

Neighboring system

- A neighboring system for a regular MRF **F** is defined as:

$$\mathbf{V} = \{Nei(F_i) \mid \forall_i \in \mathbf{F_i}\} \tag{4}$$

- **V** satisfies the following properties:
  1. A site in the field is not a neighbor to itself.
  2. The neighborhood relations are symmetric, that is, if $F_j \in Nei(F_i)$ then $F_i \in Nei(F_j)$.

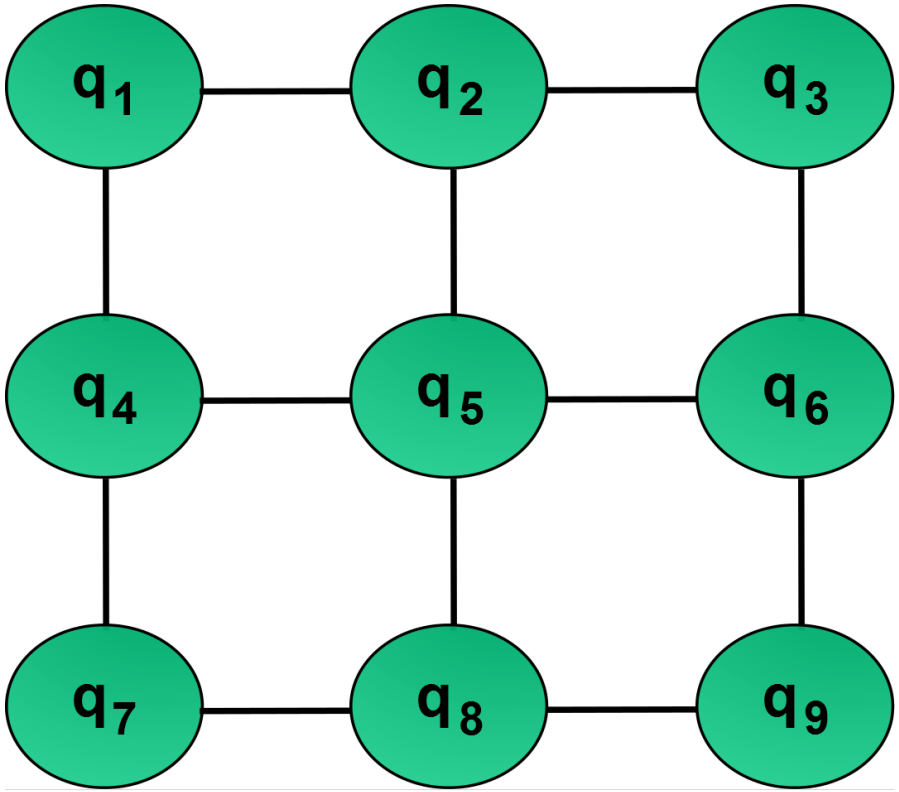
Regular Grid

- For a regular grid, a neighborhood of order i is defined as:

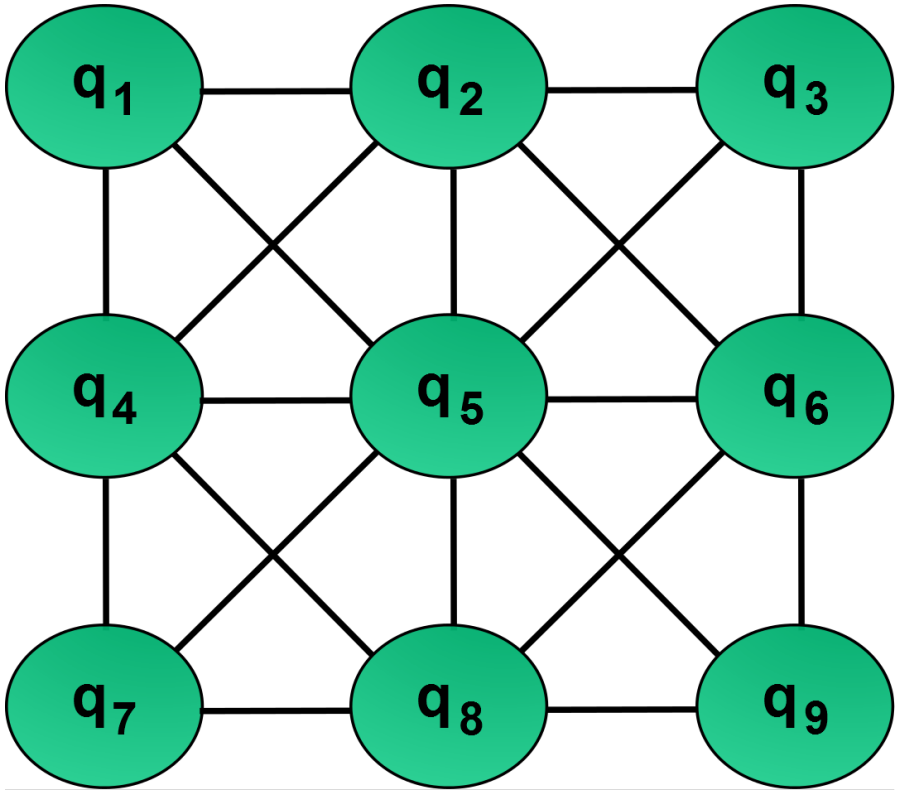
$$Nei_i(F_i) = \{F_j \in \mathbf{F} \mid dist(F_i, F_j) \leq r\} \tag{5}$$

- The radius, r, is defined for each order. For example, $r = 1$ for order one, each interior site has 4 neighbors; $r = \sqrt{2}$ for order two, each interior site has 8 neighbors; $r = 2$ for order three, each interior site has 12 neighbors

1st Order Regular Grids



2nd Order Regular Grids



Parameters

- The parameters of a regular MRF are specified by a set of local functions
- These functions correspond to joint probability distributions of subsets of completely connected variables in the graph
- In the case of a first order MRF, there are subsets of 2 variables; in the case of a second order MRF, there are subsets of 2, 3 and 4 variables
- In general:

$$P(\mathbf{F}) = (1/k) \prod_i f(X_i) \tag{6}$$

- We can think of these local functions as *constraints* that will favor certain configurations

GRM

- The joint probability of a MRF can be expressed in a more convenient way given its equivalence with a Gibbs Random Field (GRM), according to the Hammersley–Clifford theorem:

$$P(\mathbf{F}) = (1/z)\exp(-\mathbf{U}) \tag{7}$$

- **U** is known as the *energy*, given its analogy with physical energy. So maximizing $P(\mathbf{F})$ is equivalent to minimizing **U**
- The energy function can also be written in terms of local functions:

$$\mathbf{U_F} = \sum_i U_i(X_i) \tag{8}$$

Energy Function

- Considering a regular MRF of order n, the energy function can be expressed in terms of functions of subsets of completely connected variables of different sizes, $1, 2, 3, \ldots$:

$$\mathbf{U_F} = \sum_i U_1(F_i) + \sum_{i,j} U_2(F_i, F_j) + \sum_{i,j,k} U_3(F_i, F_j, f_k) + \ldots \tag{9}$$

- Given the Gibbs equivalence, the problem of finding the configuration of maximum probability for a MRF is transformed to finding the configuration of minimum energy

MRF specification

- In summary, to specify a MRF we must define:
  - A set of random variables, **F**, and their possible values, L.
  - The dependency structure, or in the case of a regular MRF a neighborhood scheme.
  - The potentials for each subset of completely connected nodes (at least the cliques).

Most Probable Configuration

- The most common application of MRFs consists in finding the most probable configuration; that is, the value for each variable that maximizes the joint probability - minimizing the energy function
- The set of all possible configurations of a MRF is usually very large, as it increases exponentially with the number of variables in **F**.
- Thus, it is impossible to calculate the energy (potential) for every configuration, except in the case of very small fields

Stochastic Search

- Usually posed as a stochastic search problem
- Starting from an initial, random assignment of each variable in the MRF, this configuration is *improved* via local operations, until a configuration of minimum energy is obtained
- After initializing all the variables with a random value, each variable is changed to an alternative value and its new energy is estimated
- If the new energy is lower than the previous one, the value is changed; otherwise, the value may also change with a certain probability

Stochastic Search Algorithm

FOR $i = 1$ TO S
$F(i) = l_k$ (Initialization)
FOR $i = 1$ TO N
FOR $j = 1$ TO S
$t = l_{k+1}$ (An alternative value for variable F(i))
IF $U(t) < U(F(i))$
$F(i) = t$ (Change value of F(i) if the energy is lower)
ELSE
IF $random(U(t) - U(F(i))) < T$
$F(i) = t$ (With certain probability change F(i) if the energy is higher)

Variants

- The way in which the *optimal* configuration is defined, for which there are two main alternatives: MAP and MPM:
- *Maximum A posteriori Probability* or MAP, the optimum configuration is taken as the configuration at the end of the iterative process
- *Maximum Posterior Marginals* or MPM, the most frequent value for each variable in all the iterations is taken as the optimum configuration

Optimization process

Iterative Conditional Modes (ICM): it always selects the configuration of minimum energy.

Metropolis: with a fixed probability, P, it selects a configuration with a higher energy.

Simulated annealing (SA): with a variable probability, $P(T)$, it selects a configuration with higher energy; where T is a parameter known as *temperature*. The probability of selecting a value with higher energy is determined based on the following expression: $P(T) = e^{-\delta U/T}$

Definition of a MRF

- The structure of the model –in the case of a regular MRF the neighborhood system.
- The form of the local probability distribution functions –for each complete set in the graph.
- The parameters of the local functions.

Estimation with labeled data

- We know the structure and functional form, and we only need to estimate the parameters
- The set of parameters, θ, of a MRF, F, are estimated from data, f, assuming no noise
- Given f, the maximum likelihood (ML) estimator maximizes the probability of the data given the parameters, $P(f \mid \theta)$; thus the optimum parameters are:

$$\theta^* = ArgMax_\theta P(f \mid \theta) \tag{10}$$

Bayesian approach

- When the prior distribution of the parameters, $P(\theta)$, is known, we can apply a Bayesian approach and maximize the posterior density obtaining the MAP estimator:

$$\theta^* = ArgMax_\theta P(\theta \mid f) \tag{11}$$

- Where:

$$P(\theta \mid f) \sim P(\theta)P(f \mid \theta) \tag{12}$$

Approximation

- The main difficulty in the ML estimation for a MRF is that it requires the evaluation of the normalizing partition function Z
- One possible approximation is based on the conditional probabilities of each variable in the field, f_i, given its neighbors, N_i: $P(f_i \mid f_{N_i})$, and assuming that these are independent - *pseudo-likelihood*
- Then the energy function can be written as:

$$U(f) = \sum_i U_i(f_i, f_{N_i}) \tag{13}$$

Assuming a first order regular MRF, only single and pairs of nodes are considered, so:

$$U_i(f_i, N_i) = V_1(f_i) + \sum_j V_2(f_i, f_j) \tag{14}$$

Pseudo-Likelihood

- The pseudo-likelihood (PL) is defined as the simple product of the conditional likelihoods:

$$PL(f) = \prod_i P(f_i \mid f_{N_i}) = \prod_i \frac{\exp -U_i(f_i, f_{N_i})}{\sum_{f_i} \exp -U_i(f_i, f_{N_i})} \quad (15)$$

- Using the PL approximation, and given a particular structure and form of the local functions, we can estimate the parameters of a MRF model based on data

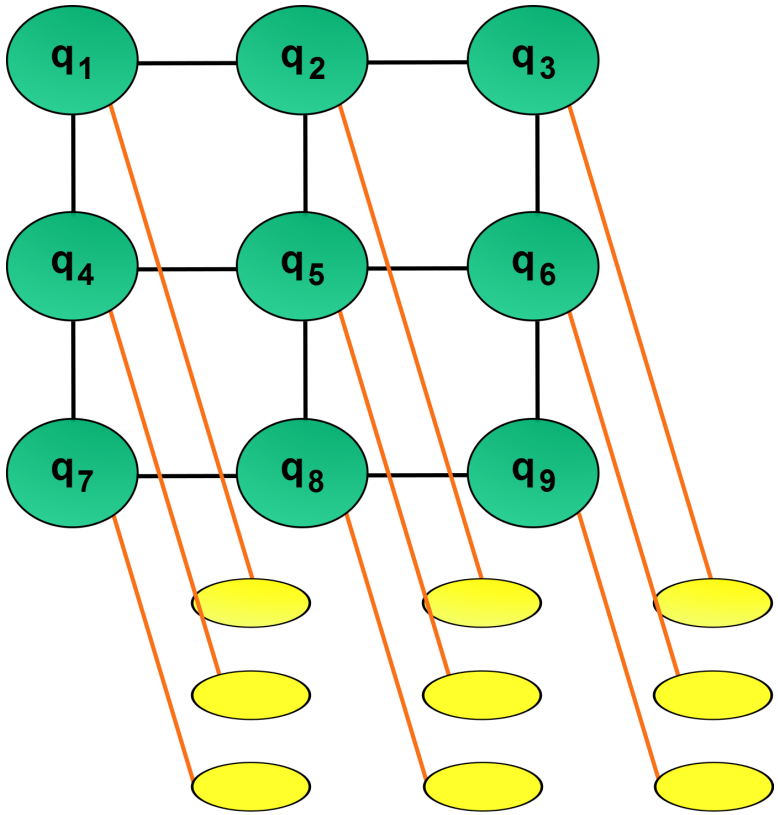
Histogram Technique

- Assuming a discrete MRF and given several realizations (examples), the parameters can be estimated using histogram techniques
- Assume there are N distinct sets of instances of size k in the dataset, and that a particular configuration (f_i, f_{N_i}) occurs H times, then an estimate of the probability of this configuration is $P(f_i, f_{N_i}) = H/N$

Image smoothing

- Digital images are usually corrupted by high frequency noise
- For reducing the noise a *smoothing* process can be applied to the image
- We can define a MRF associated to a digital image, in which each pixel corresponds to a random variable
- Considering a first order MRF, each interior variable is connected to its 4 neighbors
- Additionally, each variable is also connected to an *observation* variable that has the value of the corresponding pixel in the image

MRF for image smoothing



Restrictions

- A property of natural images is that, in general, they have certain continuity, that is, neighboring pixels will tend to have similar values
- Restrictions: (i) neighboring pixels to have similar values, by punishing (higher energy) configurations in which neighbors have different values, (ii) have a value similar to the one in the original image; so we also punish configurations in which the variables have different values to their corresponding observations
- The solution will be a compromise between these two types of restrictions

Potential functions

- The energy function can be expressed as the sum of two types of potentials: one associated to pairs of neighbors, $U_c(f_i, f_j)$; and the other for each variable and its corresponding observation, $U_o(f_i, g_i)$

$$\mathbf{U_F} = \sum_{i,j} U_c(F_i, F_j) + \lambda \sum_{i} U_o(F_i, G_i) \qquad (16)$$

- Where λ is a parameter which controls which aspect is given more importance, the observations ($\lambda > 1$) or the neighbors ($\lambda < 1$); and G_i is the observation variable associated to F_i

Potentials

- A reasonable function is the quadratic difference.
- The neighbors potential is:

$$U_c(f_i, f_j) = (f_i - f_j)^2 \tag{17}$$

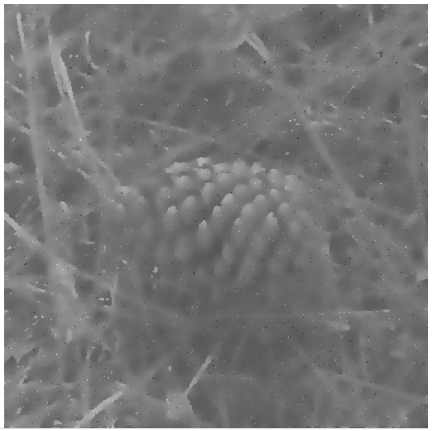
- The observation potential is:

$$U_o(f_i, g_i) = (f_i - g_i)^2 \tag{18}$$

- Using these potentials and applying the stochastic optimization algorithm, a smoothed image is obtained as the final configuration of **F**

Image smoothing





Automatic Image Annotation

- Automatic image annotation is the task of automatically assigning annotations or labels to images or segments of images, based on their local features
- When labeling a segmented image, we can incorporate additional information to improve the annotation of each region of the image
- The labels of each region of an image are usually not independent; for instance in an image of animals in the jungle, we will expect to find a sky region *above* the animal, and trees or plants *below* or *near* the animal
- *Spatial relations* between the different regions in the image can help to improve the annotation

MRFs for improving annotations

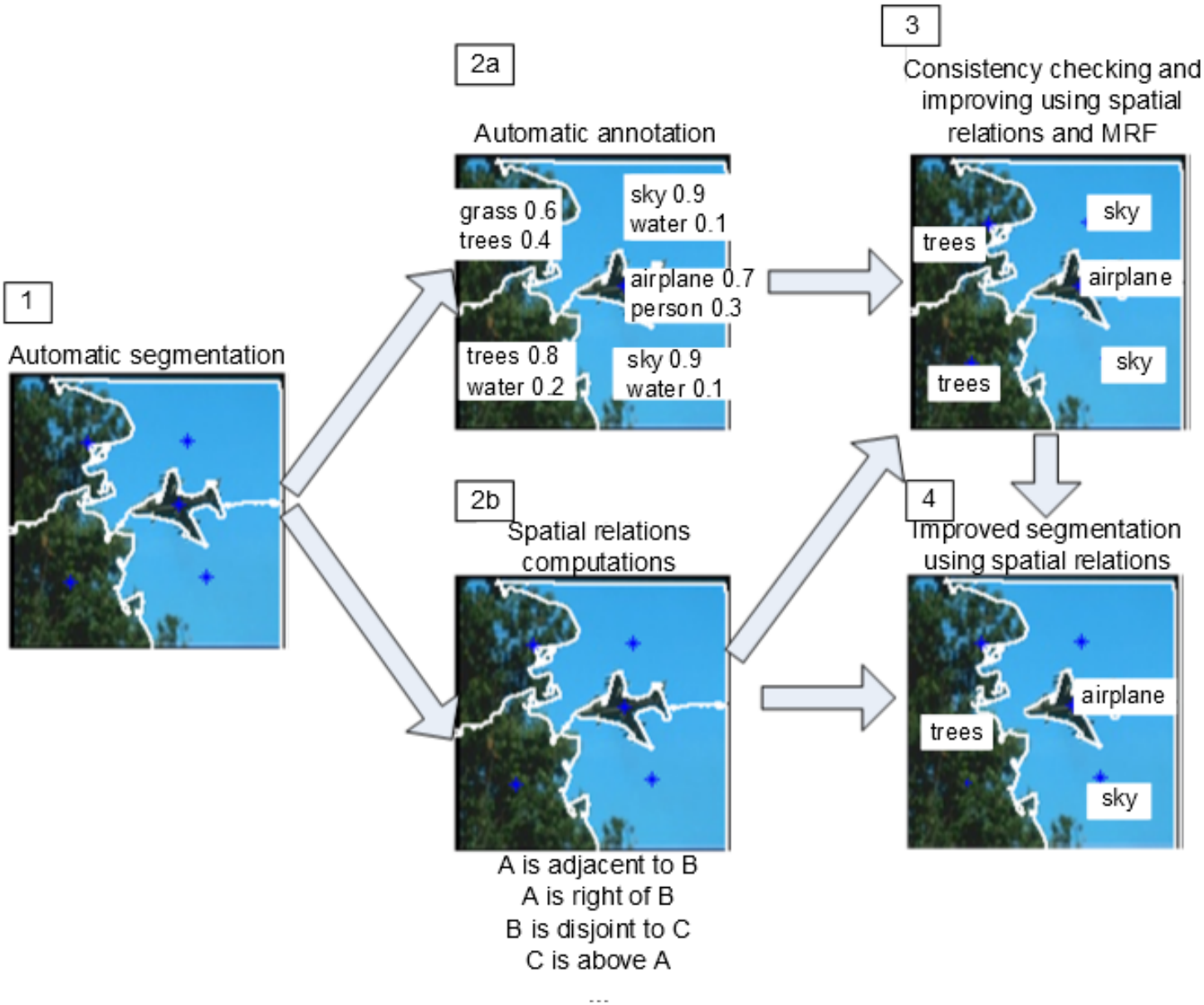
- Using a MRF we can combine the information provided by the visual features for each region (external potential) and the information from the spatial relations with other regions in the image (internal potential)
- By combining both aspects in the potential function, and applying the optimization process, we can obtain a configuration of labels that *best* describe the image

Procedure

The procedure is basically the following:

1. An image is automatically segmented (using Normalized cuts).
2. The obtained segments are assigned a list of labels and their corresponding probabilities based on their visual features using a classifier.
3. Concurrently, the spatial relations among the same regions are computed.
4. The MRF is applied, combining the original labels and the spatial relations, resulting in a new labeling for the regions by applying simulated annealing.
5. Adjacent regions with the same label are joined.

Procedure - block diagram



Energy function

- The energy function to be minimized combines the information provided by the classifiers (labels' probabilities) with the spatial relations (relations' probabilities)
- Spatial relations are divided in three groups: topological relations, horizontal relations and vertical relations - contains four terms, one for each type of spatial relation and one for the initial labels:

$$U_p(f) = \alpha_1 V_T(f) + \alpha_2 V_H(f) + \alpha_3 V_V(f) + \lambda \sum_o V_o(f) \quad (19)$$

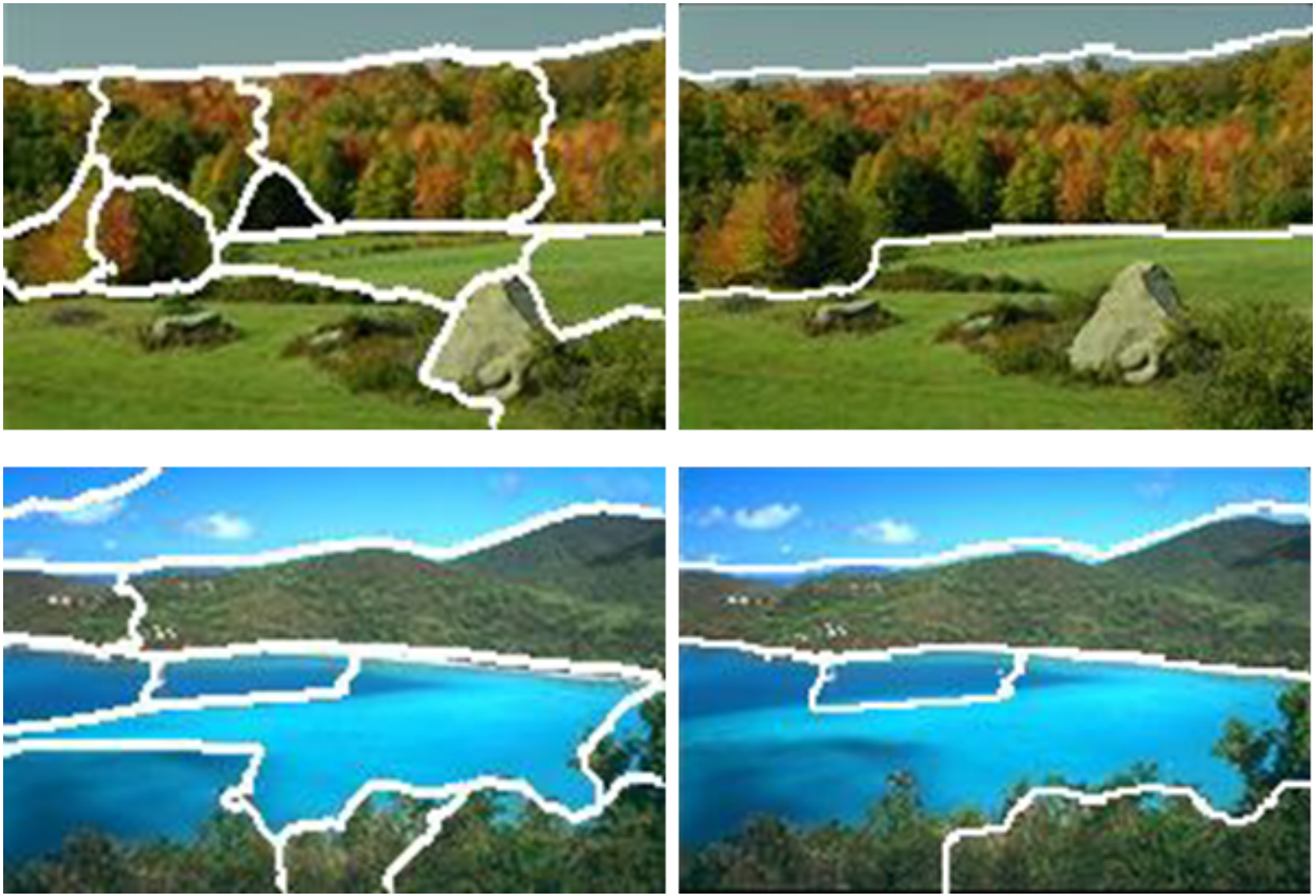
Parameters

- These potentials can be estimated from a set of labeled training images
- The potential for a certain type of spatial relation between two regions of classes A and B is inversely proportional to the probability (frequency) of that relation occurring in the training set

Application

- By applying this approach, a significant improvement can be obtained over the initial labeling of an image
- In some cases, by using the information provided by this new set of labels, we can also improve the initial image segmentation as illustrated

Example - improving segmentation



Book

Sucar, L. E, *Probabilistic Graphical Models*, Springer 2015 – Chapter 6

Additional Reading (1)


Binder, K: Ising Model. Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer-Verlag (2001).


Hammersley, J. M., Clifford, P.: Markov Fields on Finite Graphs and Lattices. Unpublished Paper. http://www.statslab.cam.ac.uk/ grg/books/hammfest/hamm-cliff.pdf (1971). Accessed 14 December 2014.


Hernández-Gracidas, C., Sucar, L.E., Montes, M.: Improving Image Retrieval by Using Spatial Relations. Journal of Multimedia Tools and Applications, vol. 62, pp. 479–505 (2013).

Additional Reading (2)

- Kindermann, R., Snell, J.L.: Markov Random Fields and Their Applications. American Mathematical Society, vol. 1 (1980).
- Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: International Conference on Machine Learning (2001).
- Li, S. Z.: Markov Random Field Modeling in Image Analysis. Springer-Verlag, London (2009).
- Sutton, C., McCallum, A.: An Introduction to Conditional Random Fields for Relational Learning. In: Geetor, L., Taskar, B. (eds.) Introduction to Statistical Relational Learning. MIT Press. (2006).