Bayesian analysis and Markov random fields for image processing

Florence Tupin
Télécom Paris

Topics

- **Image labeling**
  - ▷ Problem modeling
  - ▷ Solution with pixel independence
  - ▷ Solution with Markov Random Field
  - ▷ Exemples

- **Image restoration**
  - ▷ Problem modeling
  - ▷ Line process

- **Extensions and links with related topics**

Bayesian analysis in image processing
Data acquisition process modeling

$x \quad \rightarrow$	Degradation Detection Measure	$\rightarrow \quad y$
original scene		observation
		Pr(Y= y / X= x)

- **Space state**
  - — restoration : y_s and x_s in E (space of gray-levels)
  - — labeling : y_s in E, x_s in Λ (space of labels)

Posterior distribution

- **problem modeling :** $y \quad \rightarrow \quad x \quad ?$

$$\Pr(X = x \;/\; Y = y) = \frac{\Pr(Y = y \;/\; X = x)\;.\;\Pr(X = x)}{\Pr(Y = y)} \qquad \text{[Bayes]}$$

$\Pr(X = x \;/\; Y = y) \propto$	$\Pr(Y = y \;/\; X = x) \;.$	$\Pr(X = x)$
$\downarrow$	$\downarrow$	$\downarrow$
posterior probability	formation	prior
of x	of the observations	on the solution

- **MAP estimate :** $\hat{x} = \arg\max_{x \in \Omega} \mathbf{Pr}(X = x \;/\; Y = y)$

Punctual (per pixel) bayesian labeling

- **Example** Let us suppose a brain image labeling with 6 classes

$\Lambda = \lambda_1, \lambda_2, ..., \lambda_6$

with background, skin, bone, Gray Matter, White Matter, ventricules

- **Per pixel model**

Ech pixel is conditionally independent from its neighbors for $P(X|Y)$:

$$P(X|Y) = \Pi_s P(X_s|Y_s)$$

The problem boils down to look for the “best” label maximizing $P(X_s|Y_s)$ for each pixel s.

$$P(X_s|Y_s) \propto P(Y_s|X_s)P(X_s)$$

(per pixel MAP estimate)

Punctual bayesian labeling

◦ Likelihood

Term $P(Y_s = y_s | X_s = x_s)$

depends on the sensor (acquisition process) and considered labels.

$\Rightarrow$ physical modeling, supervised learning by manual selection of region of interest, unsupervised learning by iterative estimation (EM)

◦ Prior (per pixel)

Term $P(X_s = x_s)$

Prior knowledge on the proportion of classes

Punctual bayesian labeling

◦ **Example**

Gaussian distributions of the gray levels conditionally to the class

no prior on the class proportion

◦ **Limits**

no spatial coherency

model not adapted for image processing

$\Rightarrow$ global prior on X = Markov Random Field

Image labeling
Data acquisition process

MAP criterion $P(X = x|Y = y) \propto P(Y|X)P(X)$

- **Term $P(Y|X)$ - Hypotheses**

$$\Pr(Y = y|X = x) = \prod_{s \in S} \Pr(Y_s = y_s|x) = \prod_{s \in S} \Pr(Y_s = y_s|X_s = x_s)$$

$$P(Y|X) = \exp(-[\sum_s - \log(P(Y_s|X_s)])$$

- **Conditional probabilities $P(Y_s|X_s)$**

depend on the sensor, on the considered classes

Prior model : properties of real images (image of labels)

◦ **(if pixel independence)**

$$P(X = x) = \prod_{s \in S} P(X_s = x_s)$$

back to the per pixel bayesian classification

$P(Y_s = y_s \;/\; X_s = x_s) P(X_s) \propto P(X_s = x_s \;/\; Y_s = y_s)$

◦ **MRF hypothesis for X**

$\Rightarrow$ interaction between a pixel and its neighbors (region regularity, ...)

$$\Pr(X = x) = \frac{\exp -U(x)}{Z}$$

with $U(x) = \sum_c V_c(x)$

Posterior distribution

◦ **new Gibbs distribution**

$$\Pr(X = x \;/\; Y = y) = \frac{\exp\, -\mathcal{U}(x \;/\; y)}{Z'}$$

$$\boxed{\mathcal{U}(x \;/\; y) = \sum_{s \in S} - \log(P(Y_s = y_s \;/\; X_s)) + \sum_{c} V_c(x)}$$

$$\max_{x \in \Omega} \Pr(X = x \;/\; Y = y) \;\Leftrightarrow\; \min_{x \in \Omega} \mathcal{U}(x \;/\; y)$$

posterior field is also markovian !

Posterior distribution

◦ **Likelihood term**

$$\sum_s -\log(P(Y_s = y_s | X_s))$$

Link between the data and the label (data attachment term)

◦ **Prior term**

$$U(x) = \sum_c V_c(xs, s \in c)$$

Regularization term (does not depend on the data) to introduce prior knowledge on the searched for solution

◦ **MAP estimate**

trade-off between the data attachment term and the regularization term

Optimization

Search for the "optimal" configuration (minimizing the energy)

- **Simulated Annealing**

Gibbs distribution (for the posterior field) with decreasing temperature

Drawback : slow convergence (stochastic algorithm) but global minimum

- **ICM (Iterated Conditional Modes)**

Drawback : local minimum (deterministic algorithm) but fast convergence

Optimization

◦ ICM (Iterated Conditional Modes)

▷ Initialization $x^{(0)}$ close to the solution

▷ Sequence of images $x^{(n)}$: at step n (updating of all the sites)

— random selection of s

— state updating = max of local probabilities

$$x_s^{(n)} = \operatorname*{argmax}_{\xi \in E} P(X_s = \xi \mid y, V_s^{(n-1)})$$

▷ stop criterion : change rate $<$ threshold

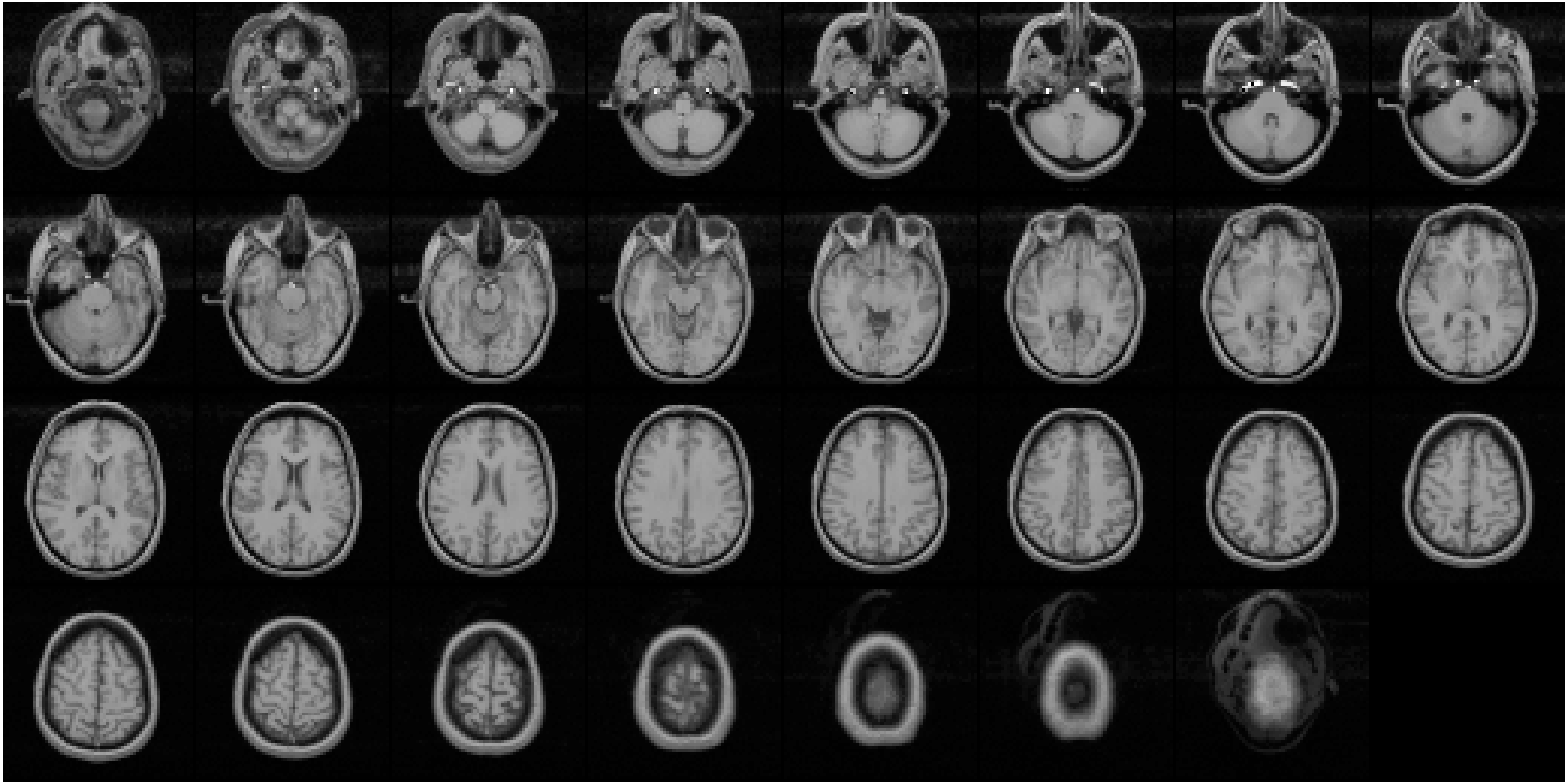
Characteristics

▷ Deterministic algorithm, result depends on initialization

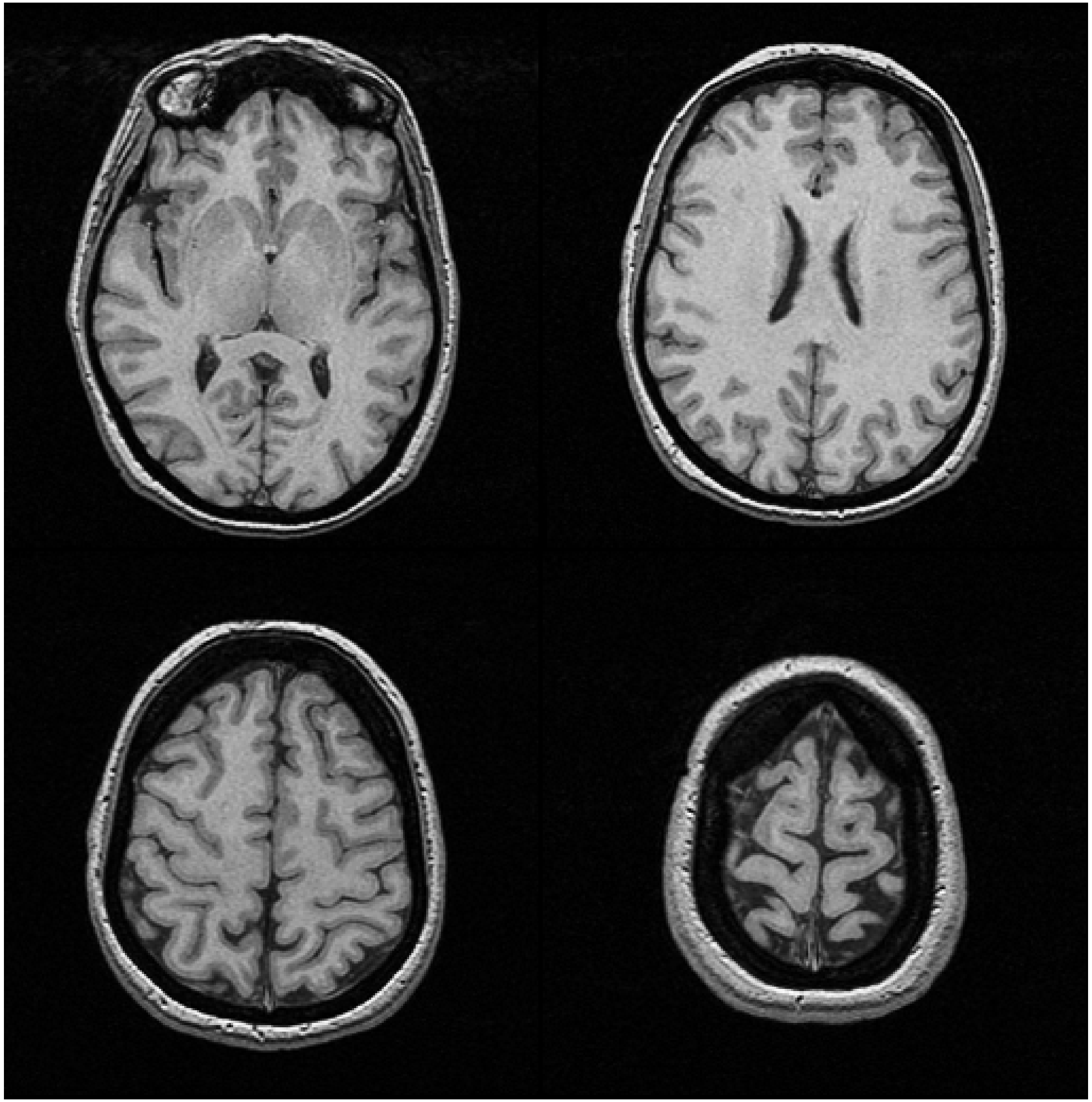
▷ Fast convergence

▷ No guarantee on the minimum of $\mathcal{U}(x \mid y)$.

Example 1



Example 1



Example 1 : brain imaging

◦ **likelihood : independence for the conditional probability**

$$P(Y = y \mid X = x) = \prod_{s \in S} P(Y_s = y_s \mid X_s = x_s)$$

Gaussian case : supervised leraning of the pdf of each class i : $\mathcal{N}(\mu_i, \sigma_i)$

$$P(Y_s = y_s \mid X_s = i) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp - \left(\frac{(y_s - \mu_i)^2}{2\sigma_i^2} \right)$$

◦ **regularisation**

Local interactions between labels : Potts model,

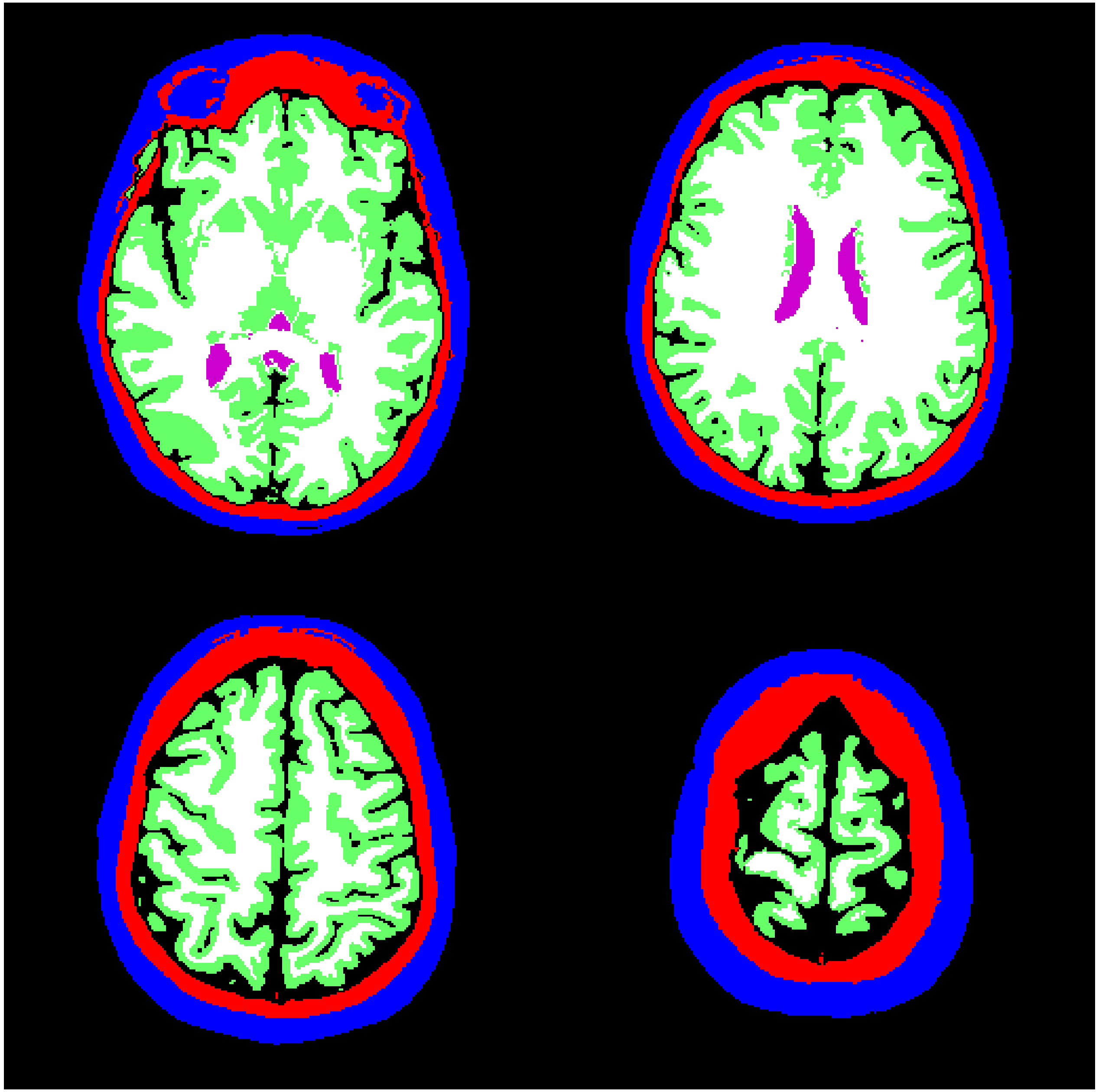
$\Rightarrow$ Posterior dist. $P(X \mid Y)$: Gibbs dist. with local conditional energy :

$$\mathcal{U}(x_s \mid y, V_s) = \log \sigma_{x_s} + \frac{(y_s - \mu_{x_s})^2}{2\sigma_{x_s}^2} + \beta \sum_{r \in \mathcal{V}_s} 1_{(x_s \neq x_r)}$$

◦ **optimization - MAP estimate $\hat{x}$**

Simulated Annealing (random init.) ; ICM (likehood estimate for initialization)

Example 1



Example 2

Example 2 : remote sensing image

◦ **Likelihood : conditional independence**

$$P(Y = y \mid X = x) = \prod_{s \in S} P(Y_s = y_s \mid X_s = x_s)$$

Gamma pdf

$$P(Y_s = y_s \mid X_s = x_s) = \frac{2L^L}{\Gamma(L)} \frac{y_s^{(2L-1)}}{\mu_{x_s}} \exp - \left(\frac{L y_s^2}{\mu_{x_s}} \right)$$

◦ **regularisation**

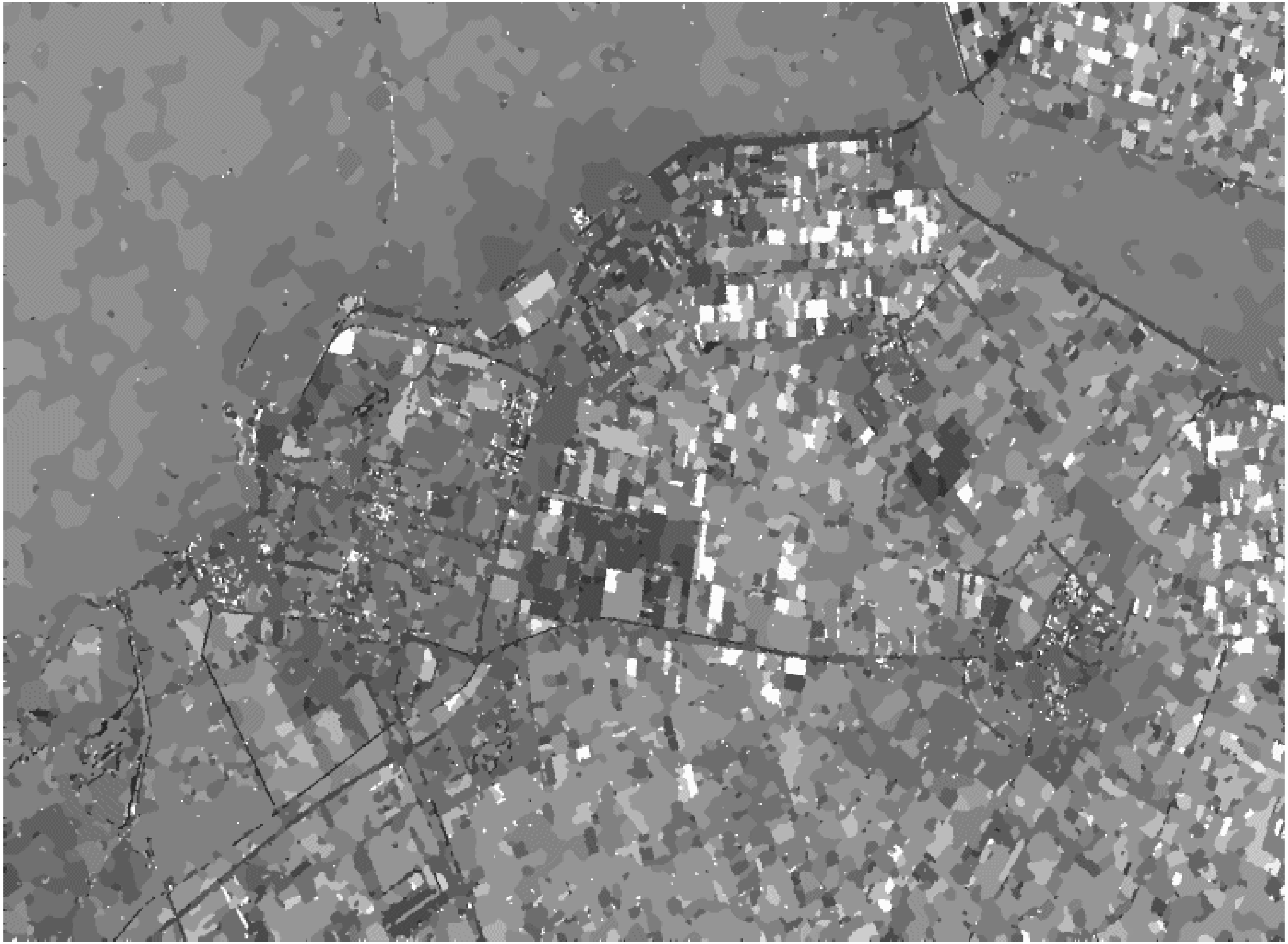
Local interactions between labels : Potts model,

◦ **Posterior : Gibbs distribution** Local conditional energy :

$$\mathcal{U}(x_s \mid y, V_s) = L \frac{y_s^2}{\mu_{x_s}} - \log \mu_{x_s} + \beta \sum_{r \in \mathcal{V}_s} 1_{(x_s \neq x_r)}$$

◦ **optimization : simulated annealing or ICM**

Exemple 2



Example 3 : segmentation and data combination

◦ **problem**

K = nomber of channels (sources) $\Rightarrow$ vector of attributes $Y = (Y^1, ..., Y^K)$

M number of classes $\Lambda = \{\lambda_1, ..., \lambda_M\}$

◦ **likelihood : independent sources**

$$
\begin{aligned}
p(Y|X) &= \prod_{s\in S} P(Y_s|X_s) = \prod_{s\in S} P(\{Y_s^1, Y_s^2, ..., Y_s^K\}|X_s) \\
&= \prod_{s\in S} P(Y_s^1|X_s)...P(Y_s^K|X_s) = \prod_{s\in S}\prod_{k=1}^{K} P(Y_s^k|X_s)
\end{aligned}
$$

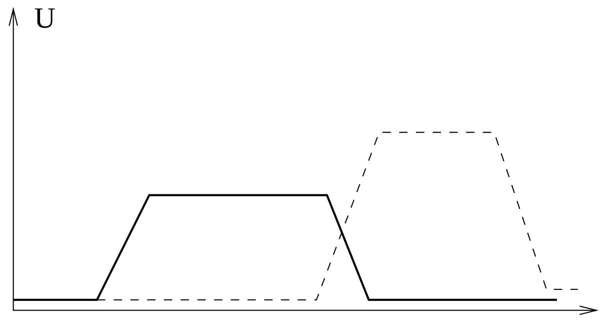
$$\Rightarrow V(y_s|\lambda) = \sum_k V(y_s^k|\lambda)$$

◦ **confidence coefficients (reliability) $C_{(k,\lambda)}$: source $k \rightarrow$ class λ**

$$V((y_s^k)_k|\lambda) = \frac{1}{\sum_k C_{(k,\lambda)}} \sum_k C_{(k,\lambda)} V(y_s^k|\lambda)$$

Segmentation and data combination

◦ **Likelihood $V(y_s^k|\lambda)$ piecewise linear**



⇒ supervised definition (histogram, thresholding,...)

⇒ automatic definition (hisogram multi-scale analysis,...)

◦ **weighting coefficients $C_{(k,\lambda)}$ for sensor k relative to λ**

= 0 if k is not significant for λ

= 0,5 if k is moderately reliable

= 1 if k is reliable for λ

Segmentation and data combination

◦ **Contextual term : Markovian label field**

$$U(x) = \sum_{c \in C} V_c(x_c)$$

◦ **Prior knowledge on class adjacency : adjacency matrix $(\gamma(\lambda_i, \lambda_j))_{i,j \in \{1,\ldots,M\}}$**

regularization potential : $V_{c=(s,t)}(x_s, x_t) = \gamma(x_s, x_t)$

⋄ forbidden adjacency between λ_1 and $\lambda_3 \Rightarrow \gamma(\lambda_1, \lambda_3) = +\infty$

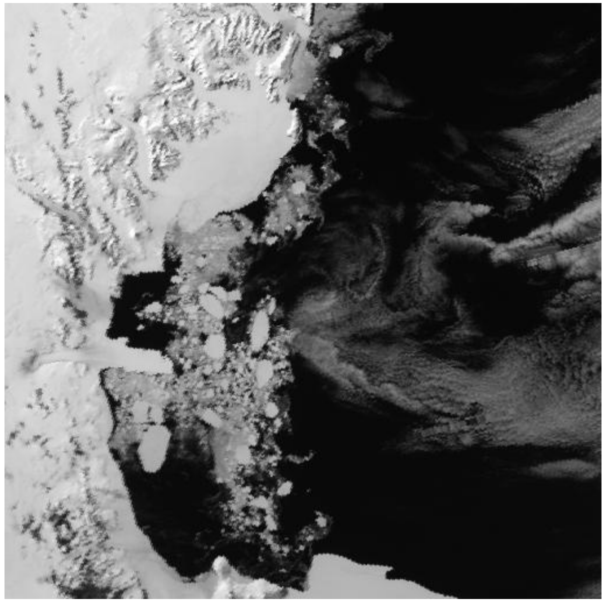
⋄ favorable adjacency for λ_1 and $\lambda_2 \Rightarrow \gamma(\lambda_1, \lambda_2) = 0$

◦ **Parameter choice**

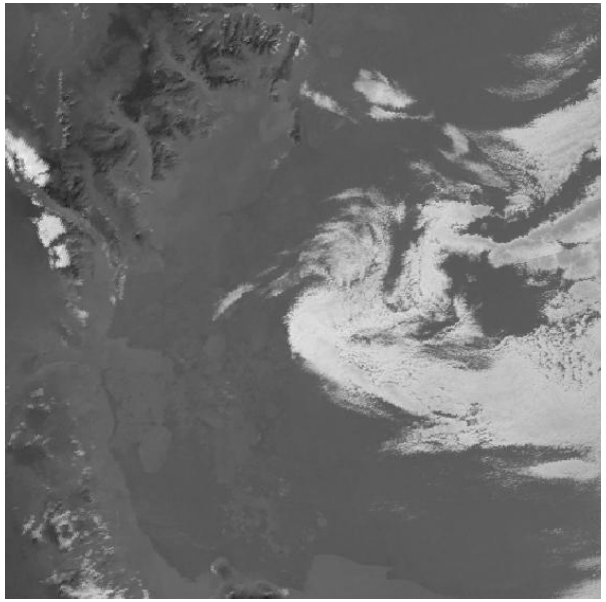
comparison of local energies for different configurations

L-curve

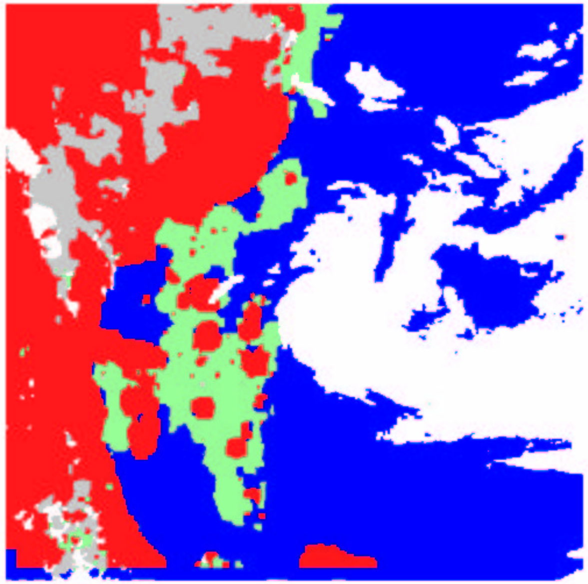
Multi-spectral labeling of AVHR RNOAA ice areas



channel 1



channel 3



labeling

$x \quad \rightarrow$	Degradation Detection Measure	$\rightarrow \quad y$
original scene		observation
		Pr(Y= y / X= x)

- **Additive white gaussian noise**

$$y = x + \epsilon \qquad y_s = x_s + \epsilon_s \; \forall s \in S \qquad \epsilon_s \rightarrow \mathcal{N}(0, \sigma^2)$$

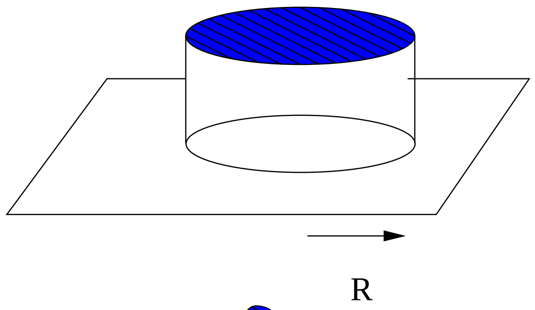
$$\Pr(Y = y \; / \; X = x) \;=\; \prod_{s \in S} \Pr(Y_s = y_s \; / \; X_s = x_s) \;\propto\; \prod_{s \in S} \exp - \frac{(y_s - x_s)^2}{2\sigma^2}$$

Loi du processus de formation des observations (suite)

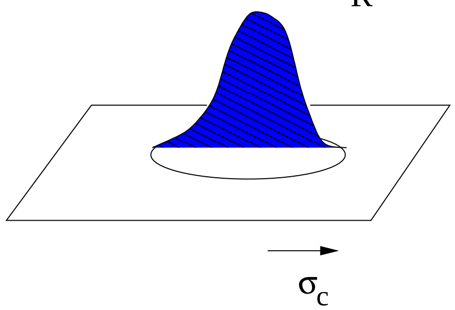
◦ **convolution**

$$\left[\; y = h\ x + \epsilon \qquad\qquad y_s = \sum_{r \in S} h_{rs}\ x_r + \epsilon_s\ \forall s \in S \qquad\qquad \epsilon_s \rightarrow \mathcal{N}(0, \sigma^2)\right.$$

- blurring (uniform)



- blurring (gaussian)



Denoising with additive white gaussian noise

$$\Pr(Y = y|X = x) \;=\; \prod_{s \in S} \Pr(Y_s = y_s|X_s = x_s) \propto \prod_{s \in S} \exp - \frac{(y_s - x_s)^2}{2\sigma^2}$$

- **regularity of solution**

$$\Pr(X = x) = \frac{\exp \; -\beta \sum_{(r,s) \in \mathcal{C}} \Phi(x_r, x_s)}{Z}$$

- **new Gibbs distribution Pr**$(X = x \;/\; Y = y) = \dfrac{\exp \; -\mathcal{U}(x \;/\; y)}{Z'}$ **!**

$$\boxed{\mathcal{U}(x \;/\; y) = \sum_{s \in S} \frac{(y_s - x_s)^2}{2\sigma^2} + \beta \sum_{(r,s) \in \mathcal{C}} \Phi(x_r, x_s)}$$

$$\max_{x \in \Omega} \Pr(X = x \;/\; Y = y) \;\Leftrightarrow\; \min_{x \in \Omega} \; \mathcal{U}(x \;/\; y)$$

- **regularization** $\Phi(x_r, x_s) = \Phi((x_r - x_s)) = \Phi(u)$

Image denoising : choice of Φ

◦ **quatratic regularization**

Gaussian field

$$\Phi(u) = u^2$$

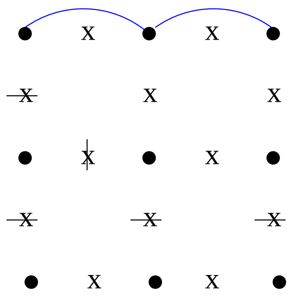
good regularization of homogeneous areas

edge blurring

◦ **suppressing the regularization term on discontinuities**

— intuitively : quadratic term $\Rightarrow$ truncated quadratic term
— introduction of a line process

Restoration taking into account discontinuities



- **Line process B**

$B = (B_{st})$

$b_{st} = 1$ if there is an edge, else $b_{st} = 0$

- **Posterior field**

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

- **Prior field energy**

$$U(x, b) = \sum_{s,t} (1 - b_{st})(x_s - x_t)^2 + \gamma b_{st}$$

Restoration taking into account discontinuities

◦ **Minimization of the energy in (x, b)**

$$\min_{(x,b)} U(x,b) = \min_x \sum_{s,t} \min_{b_{st}} f(x_s - x_t, b_{st})$$

$$\min_{b_{st}} f(x_s - x_t, b_{st}) = \min((x_s - x_t)^2, \gamma)$$

$$\min_{(x,b)} U(x,b) = \min_x \tilde{U}(x)$$

$$\min_{b_{st}} f(x_s - x_t, b_{st}) = \phi(x_s - x_t)$$

implicit model ⇔ explicit model

(*weak membrane model*)

Restoration taking into account discontinuities

◦ **examples of regularization functions $\phi(x_s - x_r)$**

Geman and Mac Clure 85 $$\phi(u) = \frac{u^2}{1+u^2}$$

Hebert and Leahy 89 $$\phi(u) = \log(1+u^2)$$

Charbonnier 94 $$\phi(u) = 2\sqrt{1+u^2} - 2$$

◦ **conditions on ϕ**

1. $$\lim_{u->0^+} \frac{\phi'(u)}{2u} = 1$$
2. $$\lim_{u->+\infty} \frac{\phi'(u)}{2u} = 0$$
3. $\frac{\phi'(u)}{2u}$ is continuous, strictly decreasing $[0, +\infty[$

Theorem

Soit :

$$\phi : [0,+\infty[\to [0,+\infty[$$

$\phi(\sqrt{u})$ strictly concave on $]0,+\infty[$
and let

$$L = \lim_{u\to+\infty} \frac{\phi'(u)}{2u} \quad \text{and} \quad M = \lim_{u\to 0^+} \frac{\phi'(u)}{2u}$$

then :

- $\exists\ \psi$ strictly convex and decreasing : $[L,M] \mapsto [\alpha,\beta]$, such that :

$$\phi(u) = \inf_{L\le b\le M} \left(bu^2 + \psi(b)\right)$$

$$\alpha = \lim_{u\to\infty} \phi(u) - u^2\frac{\phi'(u)}{2u} \quad , \quad \beta = \lim_{u\to 0^+} \phi(u) - u^2\frac{\phi'(u)}{2u}$$

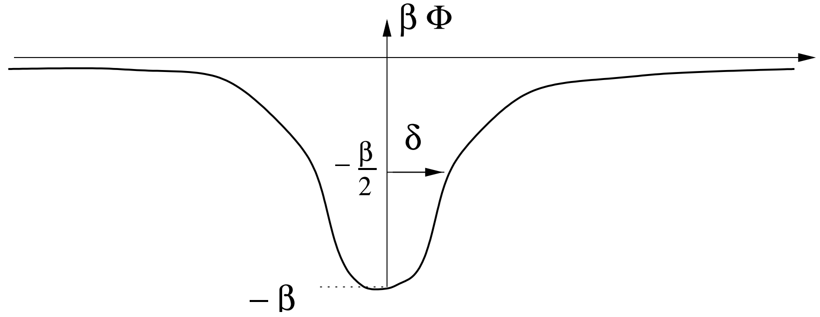
- $$\forall u \quad b_u = \frac{\phi'(u)}{2u}$$

  is the unique value for which infimum is reached

Image restoration : Geman and Reynolds potential

- **formulation**

$$\beta\ \Phi(u) = \frac{-\beta}{1 + \left(\frac{u}{\delta}\right)^2}$$



$$\begin{cases} \beta : \text{"range" of the potential} \\ \delta : \text{"Amplitude" of the potential} \end{cases}$$

- **⇒ choice of β and δ controlling the regularization**

Implicit ϕ-function vs explicit line process

to preserve discontinuitites it is strictly equivalent to minimize

- **an explicit expression with line process**

$$U(x,b|y) = \sum_s (y_s - x_s)^2 + \lambda \sum_{(r,s)} b_{rs}(x_s - x_r)^2 + \mu \sum_{(r,s)} \psi(b_{rs})$$

- **an implicit equivalent expression**

$$U(x|y) = \sum_s (y_s - x_s)^2 + \lambda' \sum_{(r,s)} \phi(x_s - x_r)$$

- **the equivalent b_{rs} is given by**

$$b_{rs} = \frac{\phi'(x_s - x_r)}{2(x_s - x_r)}$$

Minimization algorithms

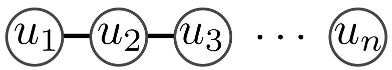
- **GNC Graduated non convexity (Blake et Zisserman)**
  - Principle : approximating the energy by a convex function and graduated modification
  - deterministic algorithm
  - proof of convergence for some cases
- **MFA Mean Field Annealing**
  - explicit line process
  - temperature decrease and mean field approximation
  - iterative estimation of the line process and the solution
- **Artur et Legend**
  - explicit line process
  - itertive computation of the line process (closed form expression) then with fixed b estimation of x (gradient descent)

MRF and graphical models

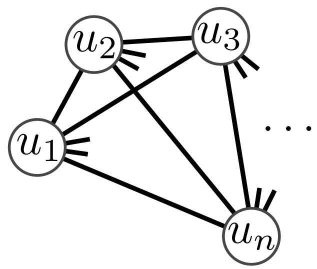
◦ Graphical models to capture independence

node = random variable, edge = probabilistic interaction

total independence



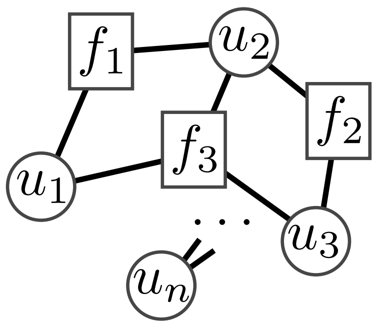
a Markov random field



complete dependence

◦ Factor graphs

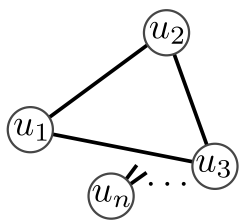
connecting groups of variables through the factor f_k



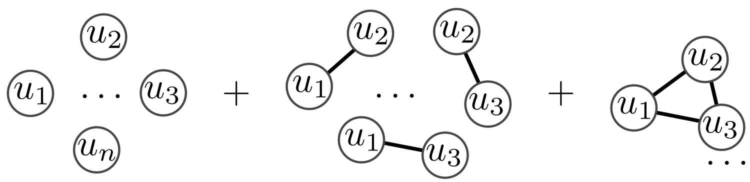
MRF and graphical models ◦ MRF

Statistical dependence of random variables and factorization

$$P(x) = \prod \psi_c(x_s, s \in c) = \frac{1}{Z} \prod_c \exp(-V_c(x))$$



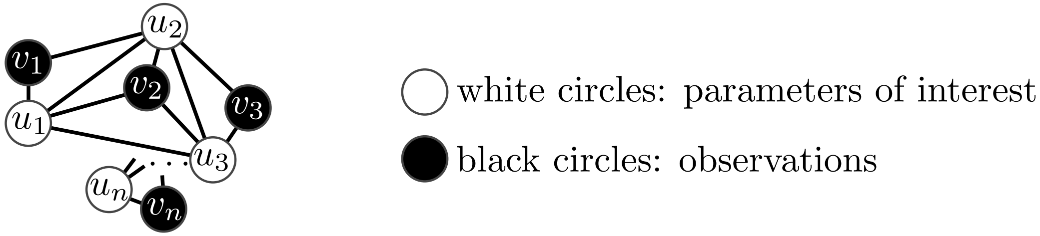
graphical model
of a Markov random field



decomposition into cliques

MRF and Conditional Random Fields (CRF)

◦ Conditional (discriminative) Random Fields



graphical model
of a conditional random field

direct modeling of the posterior field

$$P(x|y) = \frac{1}{Z} \exp(-\sum_c V_c(x, y))$$

- The clique potentials can depend on the vector of observations (external field)
- Often used in a supervised training context with a learning of $V_c(x_s, y)$ (unitary potentials) and $V_c(x_s, x_t, y)$ pairwise potentials (ex : logistic classifiers)