Analyse bayésienne et champs markoviens en traitement d'images

Florence Tupin
Télécom ParisTech

Sujets

- **Classification d'images**
  - ▷ Modélisation du problème
  - ▷ Solution sans modèle markovien
  - ▷ Solution avec modèle markovien
  - ▷ Exemples

- **Restauration d'images**
  - ▷ Modélisation du problème
  - ▷ Les processus bords

Analyse Bayesienne en Traitement des Images
Loi du processus de formation des observations

$x \quad \longrightarrow$	$\boxed{\text{Dégradation Détection Mesure}}$	$\longrightarrow \quad y$
scène originale		observation
		Pr(Y= y / X= x)

◦ **Espaces d'états**

– restauration : y_s et x_s dans E (espace des niveaux de gris)

– classification : y_s dans E, x_s dans Λ (espace des classes - *labels)*

Distribution a posteriori

- **modélise le problème :** $y \quad \longrightarrow \quad x \quad ?$

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

$$\boxed{\begin{array}{ccc} \Pr(X = x \;/\; Y = y) \propto & \Pr(Y = y \;/\; X = x) \;\; . & \Pr(X = x) \\ \downarrow & \downarrow & \downarrow \\ \text{probabilité a posteriori} & \text{formation} & \text{a priori} \\ \text{de } x & \text{des observations} & \text{sur la solution} \end{array}}$$

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

Classification d'images bayésienne ponctuelle

◦ **Exemple** Supposons qu'on veuille segmenter l'image cérébrale en 6 classes

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

avec fond, peau, os, MG, MB, ventricules

◦ **Modèle ponctuel**

On suppose que chaque pixel est indépendant de ses voisins pour $P(X|Y)$:

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

Finalement, le problème revient à chercher au pixel s la "meilleure" classe au sens du maximum de $P(X_s|Y_s)$.

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

(MAP ponctuel)

Classification d'images bayésienne ponctuelle

◦ **Vraisemblance (ponctuelle)**

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

dépend du capteur (du processus d'acquisition) et des classes considérées.

$\Rightarrow$ modélisation physique, apprentissage supervisé par sélection manuelle de régions, apprentissage automatique par méthodes itératives (EM)

◦ **A priori (ponctuel)**

Terme $P(X_s = x_s)$

Connaissance a priori de la proportion des classes

Classification d'images bayésienne ponctuelle

◦ **Exemple**

cas de distributions gaussiennes pour la distribution des niveaux de gris conditionnellement aux classes

pas d'a priori sur les classes

◦ **Limites**

pas de cohérence spatiale

modèle peu efficace en TdI

$\Rightarrow$ modèles globaux sur X = modèle markovien

Classification d'images
Loi du processus de formation des observations

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

- **Terme $P(Y|X)$ - Hypothèses**

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

- **Lois conditionnelles**

dépendent du capteur, des classes considérées

Modèle (loi) a priori : propriétés désirées sur l'image réelle (image des classes)

- **Indépendance des sites**

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

probabilité a priori d'apparition des classes

classification bayésienne "ponctuelle"

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

- **Hypothèse markovienne sur X**

$\Rightarrow$ interaction entre un site et ses voisins (régularité des régions, ...)

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

Distribution a posteriori

◦ **nouvelle distribution de Gibbs**

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

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

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

le champ *a posteriori* est markovien relativement au même système de voisinage que le champ a priori

Optimisation

Recherche de la configuration la plus probable i.e d'énergie minimale

- **Recuit simulé**

distribution de Gibbs à température décroissante

(loi a posteriori)

inconvénient : lenteur de la convergence

- **ICM (modes conditionnels itérés)**

algorithme déterministe, convergence vers un minimum local

Optimisation

◦ **ICM (modes conditionnels itérés) : algorithme déterministe**

▷ Initialisation $x^{(0)}$ proche de la solution (ex MV ponctuel)

▷ Suite d'images $x^{(n)}$: à l'étape n (balayage de tous les sites)

– tirage aléatoire d'un site s : loi uniforme ou mode raster

– nouvelle valeur = max des proba locales

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

▷ arrêt : taux de changement $<$ seuil

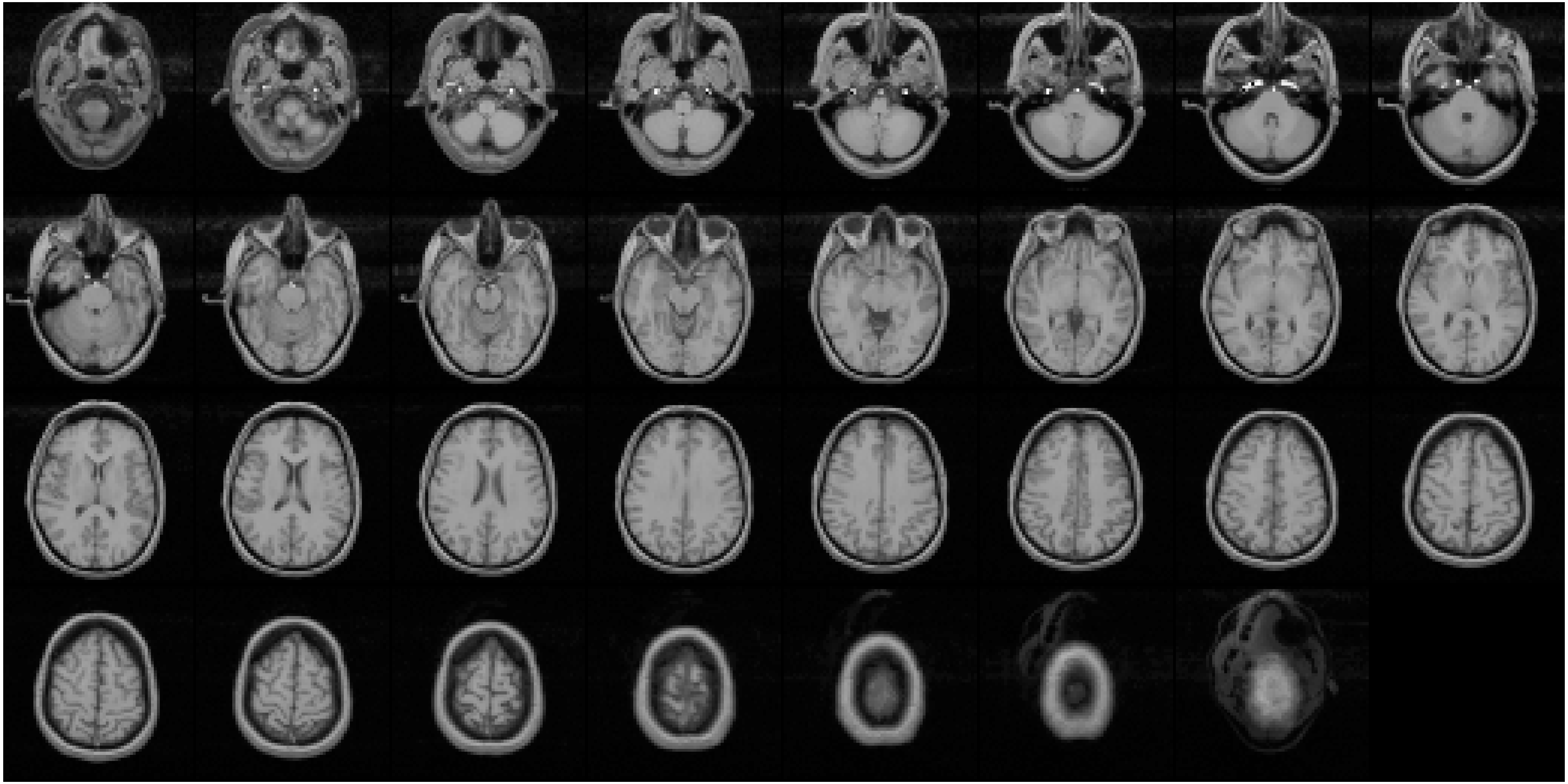
<u>Caractéristiques</u>

▷ Algorithme déterministe, résultat dépendant de l'initialisation

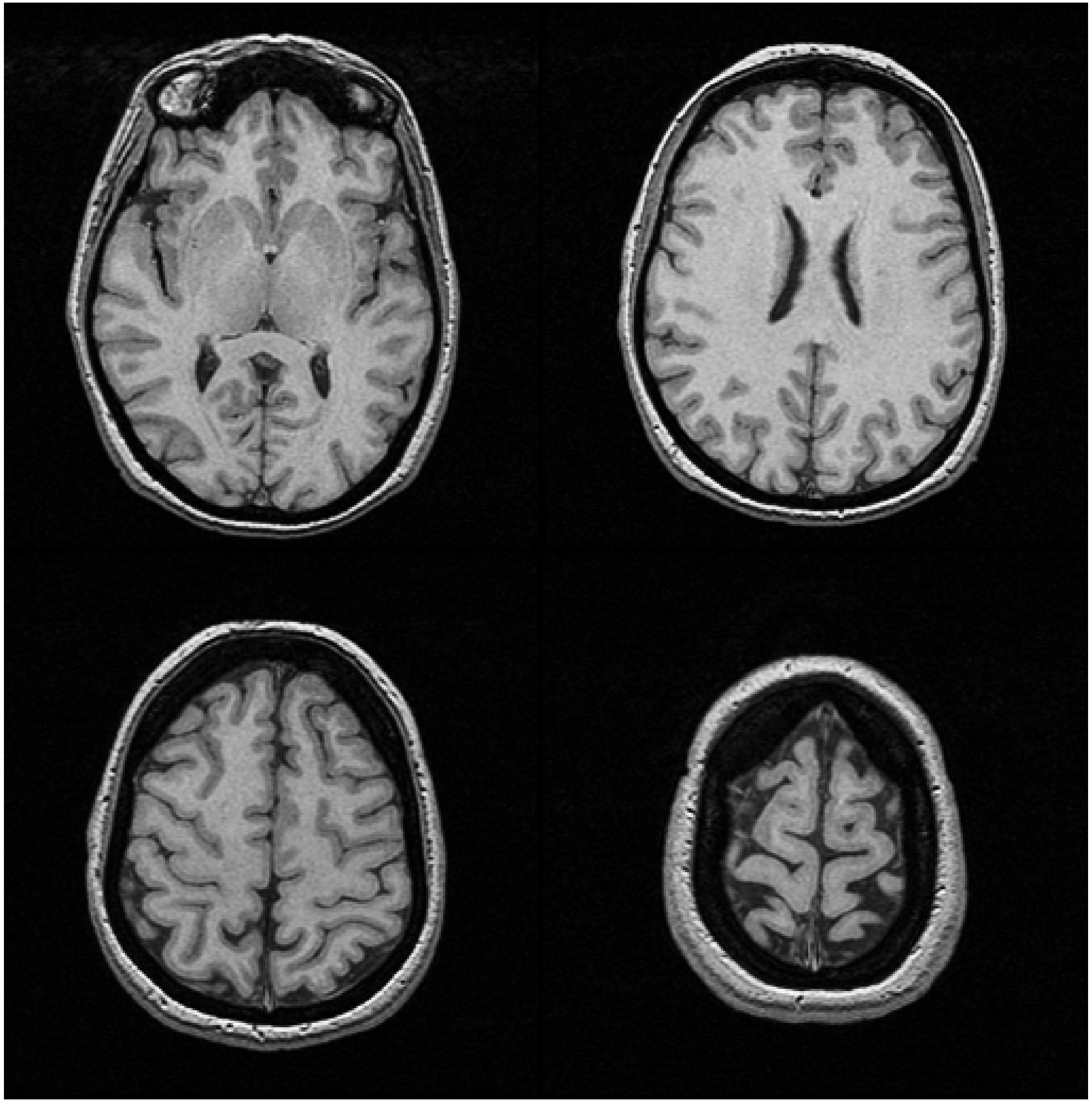
▷ Convergence rapide (quelques balayages)

▷ Risque de converger vers un minimum local de $\mathcal{U}(x \mid y)$.

Exemple 1



Exemple 1



Exemple 1 : imagerie cérébrale

- **attache aux données : indépendance des lois conditionnelles**

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

Cas gaussien : apprentissage des lois normales des classes i : $\mathcal{N}(\mu_i, \sigma_i)$

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

- **régularisation**

Intéractions locales entre étiquettes : modèle de Potts, voisinage 8-connexe

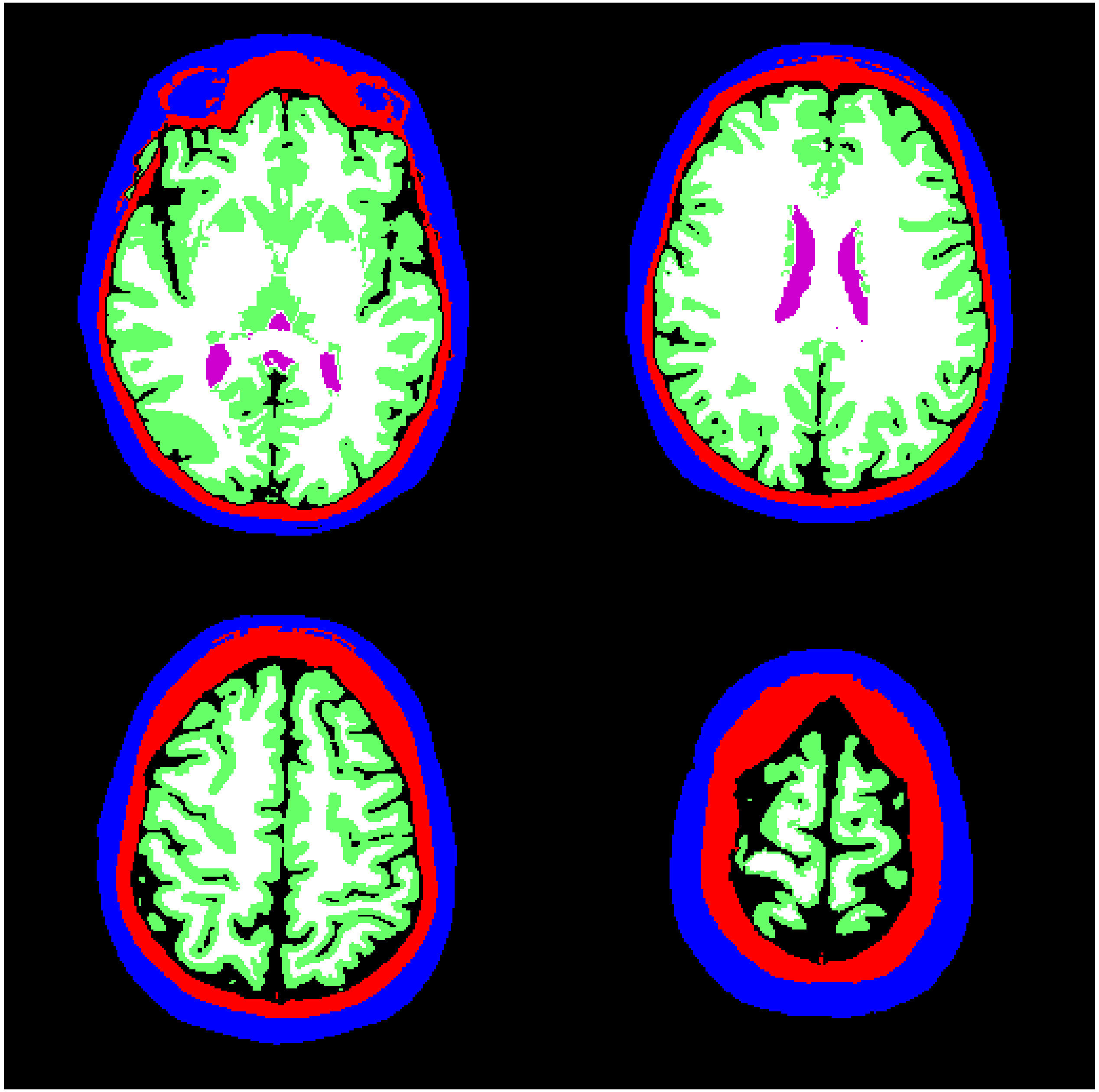
$\Rightarrow$ Distribution a posteriori $P(X \mid Y)$: loi de Gibbs d'énergie locale

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

- **optimisation - estimation au sens du MAP de x**

initialisation : aléatoire ; algorithme : recuit simulé

Exemple 1



Exemple 2

Exemple 2 : imagerie satellitaire radar

- **attache aux données : indépendance des lois conditionnelles**

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

lois Gamma

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

- **régularisation**

Intéractions locales entre étiquettes : modèle de Potts, voisinage 8-connexe

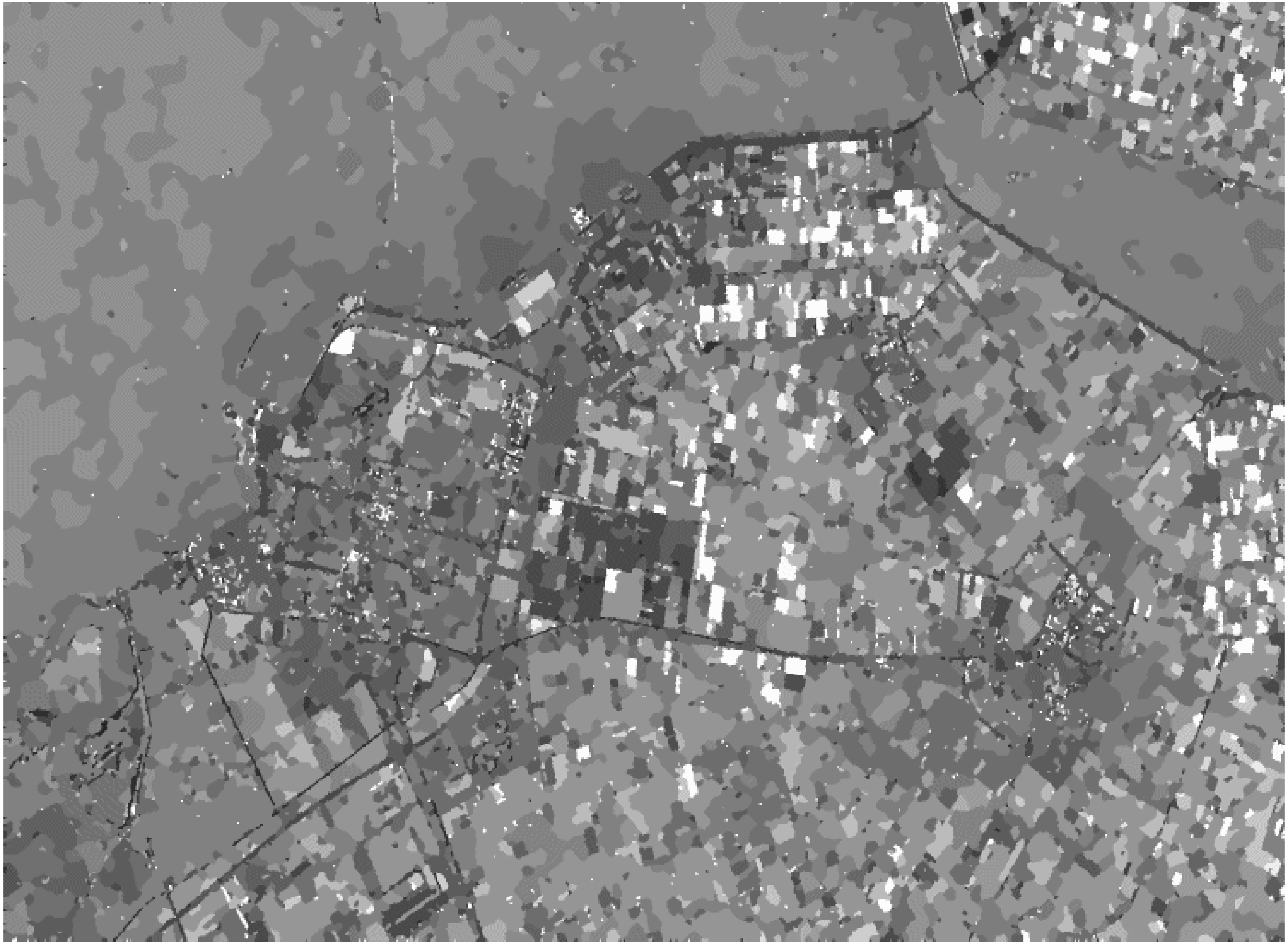
- **distribution a posteriori : loi de Gibbs**

loi de Gibbs

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

- **optimisation : recuit simulé ou ICM**

Exemple 2



Exemple 3 : segmentation et fusion de données

◦ **position du problème**

K = nombre de données (ou canaux) $\Rightarrow$ vecteur d'attributs $Y = (Y^1, ..., Y^K)$

M nombre de classes $\Lambda = \{\lambda_1, ..., \lambda_M\}$

◦ **attache aux données : indépendance des sources entre elles**

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

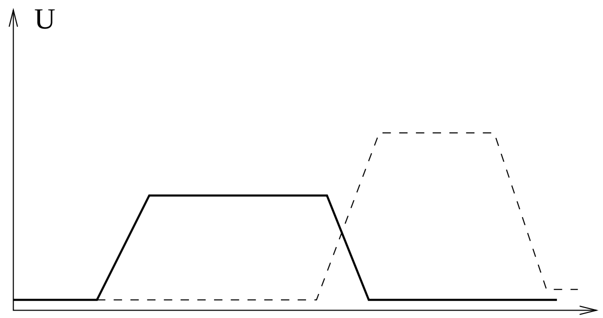
$$\Rightarrow V((y_s^k)_k|\lambda) = \sum_k V(y_s^k|\lambda)$$

◦ **coefficients de confiance (fiabilité) $C_{(k,\lambda)}$: source $k \rightarrow$ classe λ**

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

Méthode markovienne de fusion de données (suite)

◦ **potentiels d'attache aux données $V(y_s^k|\lambda)$ linéaires par morceaux**



⇒ définition supervisée (histogramme, seuillage,...)

⇒ définition automatique (analyse multi-échelles de l'histogramme, recuit simulé sous contraintes...)

◦ **coefficients de confiance $C_{(k,\lambda)}$ du capteur k relativement au label λ**

$= 0$ si source k non significative pour λ

$= 0{,}5$ si source k info. approximative

$= 1$ si source k pertinente pour λ

Méthode markovienne de fusion de données (suite)

∘ **terme contextuel : le champ des étiquettes est markovien**

$$U_{\text{context}}(x) = \sum_{c \in C} V_c(x_c)$$

∘ **connaissances a priori : matrice d'adjacence** $(\gamma(\lambda_i, \lambda_j))_{i,j \in \{1,\ldots,M\}}$

potentiel de régularisation : $V_{c=(s,t)}(x_s, x_t) = \gamma(x_s, x_t)$

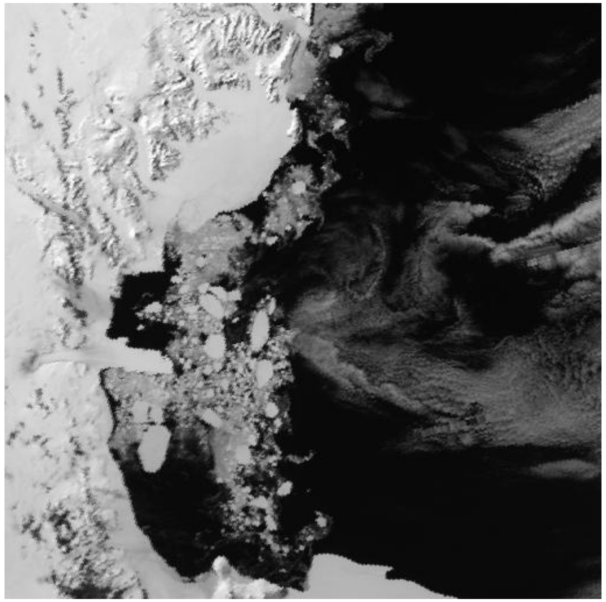
⋄ adjacence interdite entre λ_1 et $\lambda_3 \Rightarrow \gamma(\lambda_1, \lambda_3) = +\infty$

⋄ adjacence favorable entre λ_1 et $\lambda_2 \Rightarrow \gamma(\lambda_1, \lambda_2) = 0$

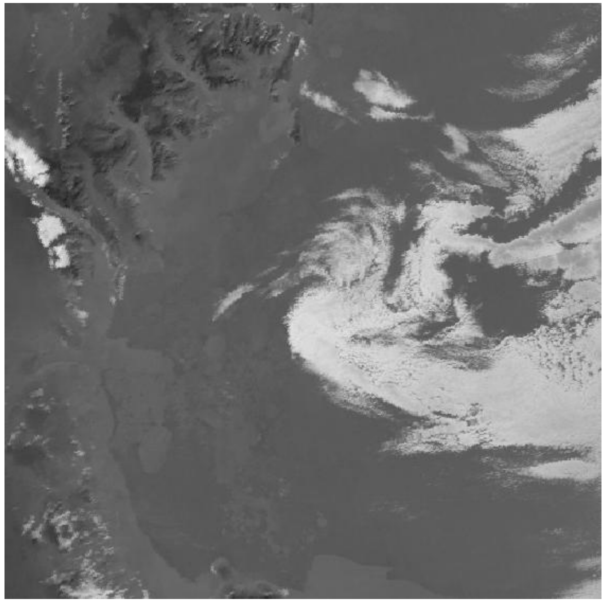
∘ **choix des paramètres** Méthode des boîtes qualitatives d'Azencott

comparaison des énergies locales dans des configurations "extrêmes"

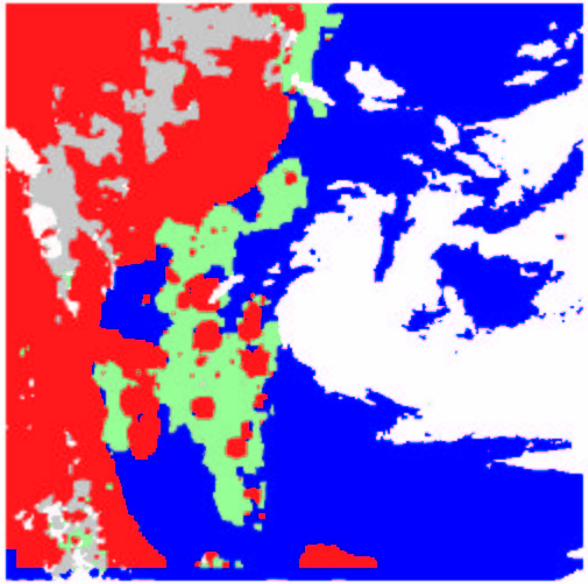
Classification multispectrale AVHR RNOAA zones glace



canal 1



canal 3



résultat

Restauration d'images
Loi du processus de formation des observations

$x \quad \rightarrow$	$\boxed{\text{Dégradation Détection Mesure}}$	$\rightarrow \quad y$
scène originale		observation
		Pr(Y= y / X= x)

◦ **bruit blanc additif gaussien**

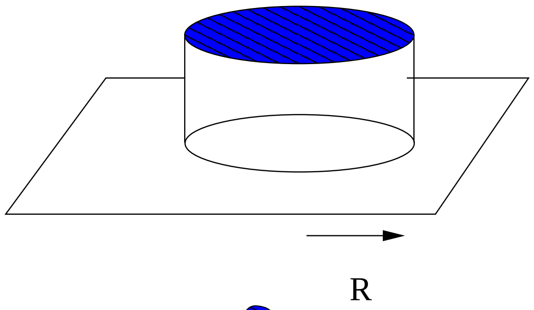
$$\left[\begin{array}{lll} y = x + \epsilon & y_s = x_s + \epsilon_s \ \forall s \in S & \epsilon_s \rightarrow \mathcal{N}(0, \sigma^2) \\ \\ \Pr(Y = y \ / \ X = x) \quad = & \prod_{s \in S} \Pr(Y_s = y_s \ / \ X_s = x_s) \quad \propto & \prod_{s \in S} \exp - \frac{(y_s - x_s)^2}{2\sigma^2} \end{array}\right.$$

Loi du processus de formation des observations (suite)

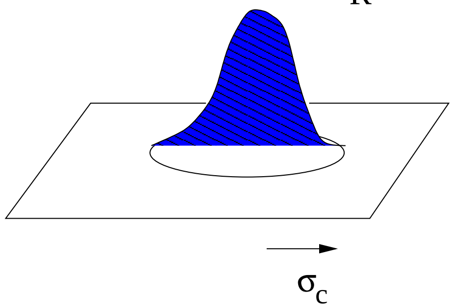
◦ **convolution**

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

- flou de bougé isotrope uniforme



- flou de mise au point (gaussien)



Débruitage d'image avec bruit blanc additif gaussien

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

◦ **régularité de l'image réelle à niveaux de gris**

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

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

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

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

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

Débruitage d'image : choix de la fonction Φ

◦ **régularisation quadratique**

champ gaussien

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

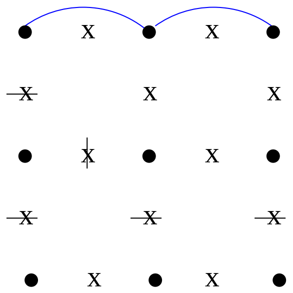
bonne régularisation des zones homogènes

dégradation des contours (flou)

◦ **suppression du lissage aux discontinuités**

– intuitivement : quadratique $\Rightarrow$ quadratique tronquée
– introduction d'un processus bords

Restauration avec prise en compte des discontinuités



- **processus bords B**

$B = (B_{st})$

$b_{st} = 1$ en présence d'une discontinuité, $b_{st} = 0$ sinon

- **champ a posteriori**

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

- **énergie a priori**

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

Restauration avec prise en compte des discontinuités

◦ **Minimisation de l'énergie en** (x, b)

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

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

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

modèle implicite $\Leftrightarrow$ modèle explicite

(modèle de la membrane mince -*weak membrane)*

Restauration avec prise en compte des discontinuités

◦ **exemples de fonctions de régularisation pour $\phi(x_s - x_r)$**

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

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

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

◦ **conditions sur les fonctions ϕ**

1. $$\lim_{u->0^+} \frac{\phi'(u)}{2u} = 1$$
2. $$\lim_{u->+\infty} \frac{\phi'(u)}{2u} = 0$$
3. $\dfrac{\phi'(u)}{2u}$ est continue et strictement décroissante sur $[0, +\infty[$

Théorème

Soit :

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

$\phi(\sqrt{u})$ strictement concave sur $]0,+\infty[$

et soit

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

Alors :

– $\exists\ \psi$ strictement convexe et décroissante : $[L, M] \mapsto [\alpha, \beta]$, tel que :

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

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

–

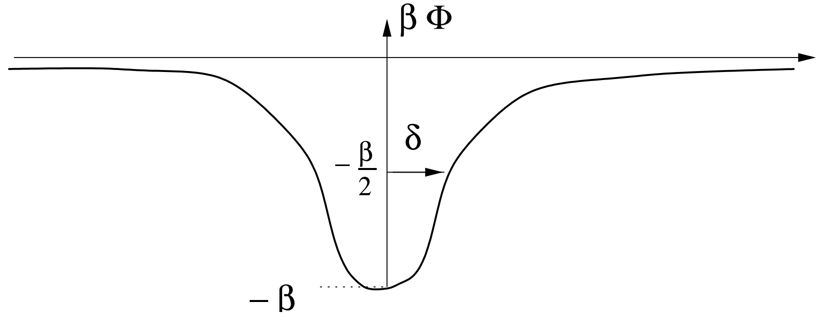
$$\forall u \quad b_u = \frac{\phi'(u)}{2u}$$

est la valeur unique pour laquelle l'infimum est atteinte

Restauration d'image : potentiel de Geman et Reynolds

◦ **sa formulation**

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



$$\begin{cases} \beta : \text{“Portée” du potentiel} \\ \\ \delta : \text{“Amplitude” du potentiel} \end{cases}$$

◦ **⇒ choix de β et δ : choix des paramètres**

Expression implicite et explicite

pour préserver les discontinuités il est strictement équivalent de minimiser

- **une expression explicite avec processus bords**

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

- **une expression implicite sans processus bords**

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

- **l'expression de b_{rs} pour l'équivalence est donnée par**

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

Algorithmes de minimisation

◦ **GNC Graduated non convexity (Blake et Zisserman)**

– Principe : approximation du critère par une fonction convexe puis modification graduelle du critère
– algorithme déterministe
– preuves de convergence seulement dans certains cas

◦ **MFA Mean Field Annealing**

– processus bords explicite
– descente de température et approximation au sens du champ moyen
– estimation itérative du processus bords et de la solution

◦ **Artur et Legend**

– processus bords explicite
– calcul itératif du processus bords (expression analytique) puis à b fixé estimation de la solution x (descente de gradient)