Segmentation d'image hyperspectrale

Serge Cohen / Gilles Celeux / Cathy Maugis

⇒ A multiscale approach for statistical characterization of functional images

Anestis Antoniadis, Xavier Brossat et Jean-Michel Poggi
Journal of Computational and Graphical Statistics (2008)

+ résultats • Kolaczyk, Ju et Gopal / Multiscale, multiresolution statistical image segmentation
Journal of the ASA (2005)

• Kolaczyk et Nowak / Multiscale likelihood analysis and complexity penalized estimation
Annals of Stat. (2004)

• Baraud, Huet, Li et al. / The MDL principle, penalized likelihoods and statistical risk, Festschrift for J. Rissanen (2008)

• Aharon, Elad et Bruckstein
K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation.
IEEE Transactions on Signal Processing (2006)

I Cadre et objectif et solution choisie

- Projet de segmentation d'image hyperspectrale
  Collaboration INRIA Saclay / Université Paris Sud / SOLEIL
  SELECT IPANEMA

IPANEMA (Institut Matériaux Anciens)

⇒ Mesure de Mesure sur des objets anciens
archéologie, paléontologie, patrimoine culturel
Inconvénients : sont de "petites" peu
⇒ Mesure de spectre très précis
d'absorption / fluorescence
⇒ Acquisition de beaucoup de spectres
⇒ Traitement automatisé → classification
non supervisé ou semi
⇒ Spécificité : ~~acquisition~~ organisation spatiale des spectres
⇒ ≥ img
⇒ Imagerie hyperspectrale
→ Imagerie satellite, Imagerie confocale
évoluée au cours du temps
(Bigot...)

Ex

Notations :

pixel : $p \in [1, N]^2 = \mathcal{P}$ $\quad N \simeq 100$

spectre : $S(p) \in \mathbb{R}^n$ ← résolution du spectre
$n \simeq 1000$

Modélisation stochastique des spectres

S "$\sim$" $\mathcal{N}(\mu, \Gamma) = P_\theta$ $\quad \mu$ moyen
Γ variance ou Γ^{-1} matrice de précision

Hypothèse : $\exists K$ classes différentes
de loi $\mathcal{N}(\mu_k, \Gamma_k) \; \forall k = P_{\theta_k}$
+ indépendance

⇒ Modèle de mélange

$$S \sim \sum_{k=1}^{K} \pi_k P_{\theta_k}$$

$$P(S=s) = \sum_{k=1}^{K} \pi_k P_{\theta_k}(S=s)$$

$$P_{\pi,\theta}(S=s) = \sum_{k=1}^{K} \pi_k P_{\theta_k}(S=s)$$

⇒ Modèle complété (avec variable cachée indiquant la classe)

$$P_{\pi,\theta}(S=s, C=k) = \pi_k P_{\theta_k}(s)$$

$$P_{\pi,\theta}(C=k) = \pi_k$$

Estimation de la "classe" de chaque pixel par MV

- Estimation de π et θ en $\hat{\pi}$ et $\hat{\theta}$

$$\hat{\pi}, \hat{\theta} = \underset{\pi, \theta}{\operatorname{argmax}} P_{\pi,\theta}(s(p)) = \underset{\pi,\theta}{\operatorname{argmax}} P_{\pi,\theta}(D)$$

$$\hat{c}(p) = \underset{\ell}{\operatorname{argmax}}\ \pi_\ell\ P_{\hat{\theta}_\ell}(s)$$

- Estimation de c, π et θ en $\hat{c}$, $\hat{\pi}$ et $\hat{\theta}$

$\hat{c}$ discrète

$$\hat{c}, \hat{\pi}, \hat{\theta} = \underset{\pi,\theta}{\operatorname{argmax}}\ \Pi\ P'_{\pi,\theta}(s[p], c[p])$$

Algorithme E.M.

Algo itératif pour calculer le M.V.

Basé sur un principe de fonction de substitution (surrogate function)

$$\log P_{\pi,\theta}(s) = \int q(c) \log P_{\pi,\theta}(s)\, dc$$

$$= \int q(c) \log \left[\frac{P_{\pi,\theta}(s,c)}{P_{\pi,\theta}(c|s)} \frac{q(c)}{q(c)} \right] dc$$

$$= -\int q(c) \log \left(\frac{q(c)}{P_{\pi,\theta}(s,c)} \right) dc$$

$$+ \int q(c) \log \left(\frac{q(c)}{P_{\pi,\theta}(c|s)} \right) dc$$

$$\log P_{\pi,\theta}(s) = -KL(q, P_{\pi,\theta}(s,c)) + KL(q, P_{\pi,\theta}(c|s))$$

Valeur courante $\hat{\pi}^e \hat{\theta}^e$ $\Rightarrow$ amélioration de la vraisemblance

$$q = P_{\hat{\pi}^e, \hat{\theta}^e}(z|s)$$

$$\Rightarrow \log P_{\pi,\theta}(s) = -KL\left(P_{\hat{\pi}^e,\hat{\theta}^e}(z|s), P_{\pi,\theta}(s,c)\right) + \underbrace{KL\left(P_{\hat{\pi}^e,\hat{\theta}^e}(c|s), P_{\pi,\theta}(c|s)\right)}_{\geq 0}$$

$$Q^e(\pi,\theta) = -KL\left(P_{\hat{\pi}^e,\hat{\theta}^e}(c|s), P_{\pi,\theta}(s,c)\right)$$

Rem
- $Q^e(\pi,\theta) \leq \log P_{\pi,\theta}(s)$
- $Q^e(\hat{\pi}^e,\hat{\theta}^e) = \log P_{\hat{\pi}^e,\hat{\theta}^e}(s)$

$\Rightarrow$ Améliorer Q^e par rapport à $Q^e(\hat{\pi}^e,\hat{\theta}^e)$ implique une amélioration sur $\log P_{\pi,\theta}(s)$

Rem
$$Q^e(\pi,\theta) = -KL\left(P_{\hat{\pi}^e,\hat{\theta}^e}(c|s), P_{\pi,\theta}(s,c)\right)$$
$$= \int P_{\hat{\pi}^e,\hat{\theta}^e}(c|s) \log P_{\pi,\theta}(s,c)\, dc - \int \hat{P}_{\pi,\theta}(c|s) \log P_{\hat{\pi},\hat{\theta}}(c|s)\, dc$$

$\Rightarrow$ Il suffit d'augmenter $\int P_{\hat{\pi}^e,\hat{\theta}^e}(c|s) \log P_{\pi,\theta}(s,c)\, dc$

Rq
$$P_{\hat{\pi}^e,\hat{\theta}^e}(c|s) = \frac{P_{\hat{\pi}^e,\hat{\theta}^e}(c,s)}{\int P_{\pi,\theta}(c,s)\, dc}$$

$\Rightarrow$ Calcul de $\int P_{\hat{\pi}^e,\hat{\theta}^e}(c|s) \log P_{\pi,\theta}(s,c)\, dc$
$\Rightarrow$ Etape E
$\Rightarrow$ Maximisation (ou augmentation) en π et θ
$\Rightarrow$ Etape M ou GM

Rq : • Algo CEM inclus ds l'étape de "classification" → complexité de l'élémentaire

• Variante stochastique pour s'échapper des minimums locaux.

• Logiciel MIXMOD

Problème en suspens,

• Nombre de classes K, structure de la matrice de covariance → pb beaucoup étudié dans le cadre de SELECT

⇒ Solution par pénalisation
⇒ Cadre de la sélection de modèles !
⇒ Cf autres exposés...

• Sélection de variables = réduction de dimension
≃ sélection de modèles "classiques" mais pb de complexité
⇒ Besoin de réduire la dimension des spectres !

• Projection sur un sous espace
⇒ ACP ⇒ variété "parcimonieuse"
→ image / Projection linéaire ⇒ variété adaptative
→ Descripteurs totalement différents

• Prise en compte de la structure spatiale
⇒ Pour l'instant absence totale !
→ Régularisation de la règle de décision
→ sur les étiquettes ⇒ Champs Markoviens (SUPELEC)
→ à travers les poids de mélange ⇒ Kolecznyk (structure multicellulaire)
Antonios ...

II Vraisemblance multi échelle, segmentation hiérarchique et estimateurs pénalisés.

Idée de Kolaczyk ~~[illegible]~~ Ju et Gopal.

Passer de $P(S(I)) = \sum \pi_\ell \, P_{\theta_\ell}(S(I))$ (i.i.d)

à $P(S(I)) = \sum \pi_\ell(I) \, P_{\theta_\ell}(S(I))$ (indépendance)

⇒ Prise en compte de la régularité spatiale à travers les poids.

⇒ Contraintes nécessaires pour rendre le modèle identifiable

→ θ_ℓ fixe !

→ π_ℓ constant par morceaux

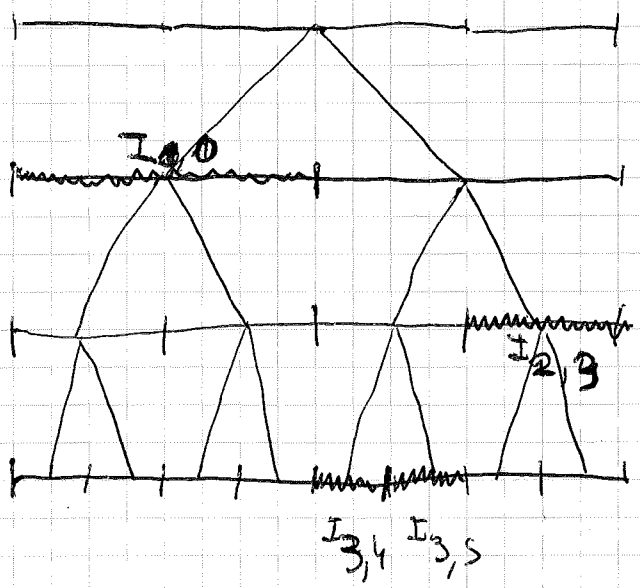
→ Partition structurée. (hiérarchique)

⇒ Cas particulier des modèles de vraisemblances multi échelles de Kolaczyk et Nowak

⇝ Algorithme et théorie !

⇒ Restriction au cas des partitions dyadiques récursives plus facile à présenter

1D



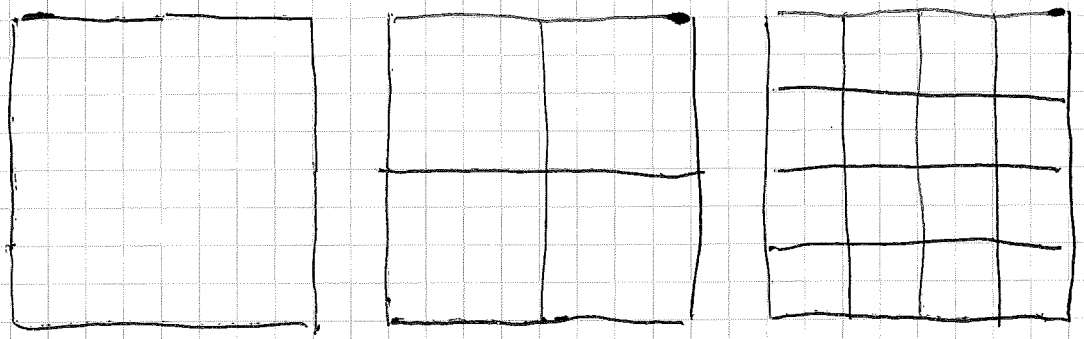
Structure d'ordre

Partition $\mathcal{P} = \{I\}$

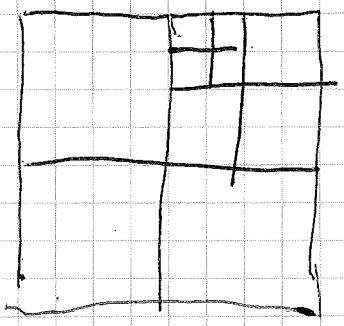
Ordre partiel

cf ondelettes

2D Même chose avec des carrés



⇒ Exemple de récursion.



Modèle : $(\mathcal{P}, (\pi_I))$ (θ_ℓ est fixé !)

Appelé MV

⇒ recherche de (π_I) optimal à $\mathcal{P}$ fixé facile

⇒ Algo "EM" sur chaque carré (calcul direct en fait)

⇒ recherche de $\mathcal{P}$ optimal ⇒ généralisation nécessaire

sur c'est toujours la partition la plus simple qui gagne !

si $\text{pen}(\mathcal{P}) = \sum_{I \in \mathcal{P}} \text{pen}(I)$, $\log P_{\mathcal{P}, \pi, \theta}(D) - \text{pen}(\mathcal{P})$

se réécrit $\sum_{I \in \mathcal{P}} \log P_{\pi_I, \theta}(S_I) - \text{pen}(I)$

⇒ séparation du problème sur chaque feuille / additivité du critère

⇒ Algorithme efficace d'optimisation de la partition

Programmation dynamique / Bottom-up / CART (Breiman)

⇒ Clé : meilleure partition sur un carré est
- soit la partition triviale
- soit l'union des meilleures partitions sur les 4 carrés qui le composent.

⇒ Algo qui commence par la partition "triviale" et qui remonte jusqu'à tester la partition "triviale" correspondant au modèle de mélange gaussien classique.

⇒ Choix de la pénalisation ?

Outil théorique : résultats de Barron et al....

Cadre général : . $D \sim P_{\gamma_0}$
. collection Γ dénombrable de γ
. $\hat{\gamma} = \operatorname{argmin}_{\gamma} \left[-\log P_\gamma(D) - \text{pen}(\gamma) \right]$

Th si $\sum_{\gamma \in \Gamma} e^{-\text{pen}(\gamma)/2} \leq 1$ alors

$$E\left(d(P_{\gamma_0}, P_{\hat{\gamma}}) \right) \leq \inf_{\gamma} \left[KL(P_{\gamma_0}, P_\gamma) + \text{pen}(\gamma) \right]$$

Explication

$$KL(P_{\gamma_0}, P_\gamma) = E_{P_{\gamma_0}} \left[-\log \frac{dP_\gamma}{dP_{\gamma_0}} \right] \quad \text{Kullback}$$

$$A(P_{\gamma_0}, P_\gamma) = E_{P_{\gamma_0}} \left[\left(\frac{dP_\gamma}{dP_{\gamma_0}} \right)^{1/2} \right] = \int \left(\frac{dP_\gamma}{d\lambda} \right)^{1/2} \left(\frac{dP_{\gamma_0}}{d\lambda} \right)^{1/2} d\lambda$$

$$h^2(P_{\gamma_0}, P_\gamma) = \int \left(\left(\frac{dP_{\gamma_0}}{d\lambda} \right)^{1/2} - \left(\frac{dP_\gamma}{d\lambda} \right)^{1/2} \right)^2 d\lambda = 2 - 2A(P_{\gamma_0}, P_\gamma)$$
$$\left(A = 1 - \tfrac{1}{2} h^2 \right)$$

$$d(P_{\gamma_0}, P_\gamma) = -2 \log A(P_{\gamma_0}, P_\gamma)$$

Ordre $\quad h^2(P_{\gamma_0}, P_\gamma) \leq d(P_{\gamma_0}, P_\gamma) \leq KL(P_{\gamma_0}, P_\gamma)$

1 égalité oracle

Rq : si $\| \log \frac{dP_\gamma(D)}{dP_{\gamma_0}(D)} \|_\infty \leq B$ alors

$$d(P_{\gamma_0}, P_\gamma) \leq KL(P_{\gamma_0}, P_\gamma) \leq (2+B)\, d(P_{\gamma_0}, P_\gamma)$$

Preuve :

$$d(P_{\gamma_0}, P_{\hat\gamma}) = -2 \log A(P_{\gamma_0}, P_{\hat\gamma})$$

$$= -2 \log A(P_{\gamma_0}, P_{\hat\gamma}) + \log \frac{dP_{\gamma_0}(D)}{dP_{\hat\gamma}(D)} + pen(\hat\gamma)$$

$$+ 2 \log \left(\frac{dP_{\hat\gamma}}{dP_{\gamma_0}}\right)^{1/2} + 2 \log\left(e^{-\frac{pen(\hat\gamma)}{2}}\right)$$

$$= 2 \log \left[\frac{\left(\frac{dP_{\hat\gamma}}{dP_{\gamma_0}}\right)^{1/2} e^{-\frac{pen(\hat\gamma)}{2}}}{E_{P_{\gamma_0}}\left(\left(\frac{dP_{\hat\gamma}}{dP_{\gamma_0}}\right)^{1/2}\right)} \right] + \log \frac{dP_{\gamma_0}(D)}{dP_{\hat\gamma}(D)} + pen(\hat\gamma)$$

$$\leq 2 \log \left[\sum_{\gamma \in \Gamma} \frac{\left(\frac{dP_\gamma}{dP_{\gamma_0}}\right)^{1/2} e^{-pen(\gamma)}}{E_{P_{\gamma_0}}\left(\left(\frac{dP_\gamma}{dP_{\gamma_0}}\right)^{1/2}\right)} \right] + \min_{\gamma \in \Gamma} \left[-\log \frac{dP_{\gamma_0}}{dP_\gamma} + pen(\gamma) \right]$$

$$E\left(d(P_{\gamma_0}, P_{\hat\gamma})\right) \leq 2 E \log \left[\sum_\gamma \quad \right] + E \min_{\gamma \in \Gamma}$$

$$\leq 2 \log E\left(\sum_\gamma [\,]\right) + \min_{\gamma \in \Gamma} E\left(\quad\right)$$

$$\leq \underbrace{2 \log \left(\sum_\gamma e^{-pen(\gamma)}\right)}_{\leq 0} + \min_{\gamma \in \Gamma} \left(KL(P_{\gamma_0}, P_\gamma) + pen(\gamma) \right)$$

Ici : il faut discrétiser le problème

$\Rightarrow$ partition discrète par nature

$\Rightarrow$ π_I restreint à une grille de précision $N^{-\beta}$

Rq $\sum_\delta e^{-\frac{pen(\delta)}{2}} \leq 1$ $\Rightarrow$ Kraft en théorie de l'information

$\Rightarrow$ $\frac{pen(\delta)}{2}$ $\simeq$ coût de codage de la loi

Ici : • $\mathcal{P}$ partition à $|\mathcal{P}|$ éléments.

$\Rightarrow$ codage à travers l'arbre

$\Rightarrow$ $|\mathcal{P}|$ feuilles $\Rightarrow$ $\frac{|\mathcal{P}|-1}{3}$ nœuds intérieurs

(chaque nœuds intérieur à 4 fils ....)

$\Rightarrow$ Codage avec $|\mathcal{P}|$ 1 et $\frac{|\mathcal{P}|-1}{3}$ 0

$\Rightarrow$ Coût $\left[\frac{4}{3}|\mathcal{P}| - \frac{1}{3}\right] \log 2$

• Dans chaque carré, spécification de π_I : $k-1$ nombres dans $[0,1]$ avec une précision $N^{-\beta}$

$\Rightarrow$ $(k-1)\beta \log N$ nats ($\leftarrow$ bits en base e)

$\Rightarrow$ $pen(\mathcal{P}, \pi_I) = 2 \times \left[\left[\frac{4}{3}|\mathcal{P}| - \frac{1}{3}\right]\log 2 + |\mathcal{P}|(k-1)\beta \log N\right]$

$= \left[\frac{8}{3}\log 2 + 2(k-1)\beta \log N\right]|\mathcal{P}| - \frac{2}{3}\log 2$

$= \sum_{I \in \mathcal{P}} \underbrace{\left(\frac{8}{3}\log 2 + 2(k-1)\beta \log N\right)}_{pen(I)} - \frac{2}{3}\log 2$

Rq : raffinement possible si on fait varier le nombre de classe

Contrôle de $KL(P_{\gamma_0}, P_\gamma) + pen(\gamma)$

Ici $P_{\gamma_0} : P(s(\rho)) = \sum_\ell \pi^0_\ell(\rho)\, P_{\theta^0_\ell}(s(\rho))$

contre $P_\gamma : P(s(\rho)) = \sum_\ell \pi_{I(\rho)}\, P_{\theta^0_\ell}(s(\rho))$

$$\Rightarrow KL(P_{\gamma_0}, P_\gamma) \le 2 \sum_\rho \sum_\ell \left| \pi^0_\ell(\rho) - \pi_{I(\rho)} \right|$$

$$\min_\pi KL(P_{\gamma_0}, P_\gamma) \le 2 K \| \pi^0(\rho) - \pi^0_\beta(\rho) \|_1 + 2K N^2 \cdot N^{-\beta}$$

Théorie de l'approximation sur π^0

$$\min_{|\rho| \le M} \frac{1}{N^2} \| \pi^0(\rho) - \pi^0_\beta(\rho) \|_1 \le C K M^{-\alpha}$$

$$\frac{1}{N^2} \left(\min_\beta \min_\pi KL(P_{\gamma_0}, P_\gamma) + pen(\gamma) \right) \le 2CKM^{-\alpha} + 2KN^{-\beta} + (C_1 + C_2 \beta \log N) \frac{M}{N^2}$$

$\Rightarrow$ Optimisation en M et $\beta = 1$

$$\le C \left(\frac{\log N}{N^2} \right)^{\frac{\alpha}{\alpha+1}}$$

$\Rightarrow$ Vitesse "habituelle" en dimension 2 !

$\Rightarrow$ Extension possible pour éviter la discrétisation ?

$\rightsquigarrow$ passage par une ligne superélisation

$\rightsquigarrow$ argument d'entropie (cf preuve des vitesses de modèles)

$\Rightarrow$ Bootstrap possible pour resserrer les Θ_ℓ !

III Réduction de dimension initiale

→ Algo automatique limité par la dimension du problème
→ Pénalisation ℓ^1 sur la matrice de précision ⇒ Chistole...
⇒ Méthode ad hoc de réduction de dimension

- Antoniadis, Bigot, von Sachs : autre méthode basée sur les ondelettes
  ⇒ seuillage des coeffs d'ondelettes
  ⇒ Conservation de l'ensemble des coeffs significatifs sur les courbes
  ⇒ Elimination des coefficients qui semblent constants...
  → Justification théorique !

- Méthode basée sur l'apprentissage de la représentation
  ex K-SVD de Elad et al.
  ⇒ Recherche de $S_1 \dots S_d$ tel que toutes les courbes observées se représentent correctement de manière parcimonieuse

$$\underset{S_1 \dots S_d}{\operatorname{argmin}} \sum_p \min_{\|\alpha\|_0 \le C} \|S[p] - \sum \alpha_i S_i\|^2$$

Cas $C = 1$ → K-Means ≃ clustering
≃ Recherche d'espace affine adapté à chaque cluster

Clustering sur les p.s. $c_i[p] < S[p], S_d >$

→ résultats préliminaires encourageants !

- Comment combiner cette recherche de représentation avec l'objectif de clustering ??
  Déjà fait pour cadre supervisé mais adaptation à un supervisé non trivial !!!