

Segmentation bayésienne d’images constantes par morceaux cas de distorsion non-linéaire, données manquantes et bruit de Poisson

Jean-François GIOVANNELLI¹, Yoann ALTMANN²

¹IMS (Univ.Bordeaux, CNRS, BINP), F-33400 Talence, France

²School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK

Giova@IMS-Bordeaux.fr, Y.Altmann@hw.ac.uk

Résumé – Cet article explore le problème de segmentation d’images constantes par morceaux, contaminées par un bruit de Poisson. Le travail inclut deux types de dégradations supplémentaires: pixels manquant et gain non-linéaire. Notre solution suit une approche bayésienne, ce qui permet d’assurer estimation et segmentation optimales. Un échantillonneur de Gibbs, incluant des mouvements de Metropolis-Hastings explore la loi a posteriori. Les échantillons sont alors utilisés pour approcher les quantités d’intérêt, en particulier amplitudes des classes et étiquettes de segmentation. Malgré la présence des dégradations, les résultats sont encourageants et généralement proches de ceux obtenus sans dégradation.

Abstract – The paper investigates the problem of segmentation of piecewise constant image corrupted by Poisson noise. A double degradation is accounted for: missing data and non-linear gain. The proposed solution follows a Bayesian strategy and a numerical approach is used to explore the intricate posterior distribution: a Gibbs sampler including a Metropolis-Hastings step. The samples generated are subsequently used to compute the amplitudes and the segmented image. The first numerical evaluation provides encouraging results despite the double degradation.

1 Introduction

La segmentation d’images est un problème complexe déjà traité d’un très grand nombre de manière [1, 2]. Ce travail, se concentre sur un cas particulier pour lequel (à notre connaissance) aucune solution statistiquement robuste n’a été proposée : la segmentation d’images constantes par morceaux [3] contaminées par 1) un bruit de Poisson, 2) des pixels manquantes et 3) une distorsion non-linéaire affectant les amplitudes. Ce contexte est associé à des procédés de détections de particules et des processus de comptage (*e.g.*, photons ou neutrons) dans un régime de faible flux (< 50 détections par pixel). De plus, la distorsion modélise de possibles imperfections (saturation) des capteurs ou des sources de particules.

On se place dans un cadre bayésien et on définit un modèle hiérarchique pour les inconnues (amplitudes des classes, étiquettes des pixels et hyperparamètres). En particulier, un champ de Potts inclut la cohérence spatiale des étiquettes. La loi a posteriori est explorée par des méthodes de simulation stochastique, Gibbs et Metropolis-Hastings (MH). Bien que des problèmes similaires (bruit gaussien) aient été traités, l’originalité du travail réside également dans la construction et la comparaison de diverses propositions pour les étapes de MH, afin d’améliorer les propriétés de mélange et de vitesse de convergence.

Les avantages principaux de la contribution sont 1) l’optimalité de l’estimation et de la décision (au sens du risque), 2) la flexibilité (facilité de généralisation à d’autres modèles d’observation, *e.g.*, processus de comptage et de distorsion), et 3) le faible niveau de supervision (le seul paramètre crucial à régler, si inconnu, est le nombre de classes). En effet, ici, le paramètre du champs de Potts est estimé conjointement avec les autres.

2 Modélisation et estimation

2.1 Image et système d’observation

On s’intéresse aux images constantes par classe et leur modélisation repose essentiellement sur deux jeux de variables :

1. pour $p = 1, \dots, P$, des étiquettes $\ell_p \in \{1, \dots, K\}$ et
2. pour $k = 1, \dots, K$, des amplitudes $\lambda_k \in \mathbb{R}$,

en notant P et K les nombres de pixels et de classes. Pour alléger les notations, ℓ et λ collectionnent $\{\ell_p\}_p$ et $\{\lambda_k\}_k$.

Afin de formaliser leur arrangement en une image constante par classe, on définit une indicatrice i_k pour chaque k comme suit. Pour une configuration ℓ des étiquettes, $i_k(\ell)$ est un vecteur binaire de taille P dont la p -ième composante vaut $\delta(\ell_p; k)$, où δ est le symbole de Kronecker. L’arrangement en une image x constante par classe, sous forme vectorielle, s’écrit alors :

$$x = \sum_k \lambda_k i_k(\ell). \quad (1)$$

En d’autres termes : le pixel p a une amplitude $x_p = \lambda_{\ell_p}$. Par ailleurs, les observations incluent deux dégradations :

1. Distorsion : chaque amplitude subit une distorsion non-linéaire notée φ , modélisant par exemple une saturation comme $\varphi(u) = s \tanh(\alpha u/s)$ où s représente un niveau de saturation et α un gain.
2. Données manquantes : seulement $M \leq P$ pixels sont observés sur l’ensemble de P pixels.

Pour terminer, les données résultent d’un processus de comptage : pour chaque pixel observé p , on acquiert indépendamment une réalisation y_p d’une variable de Poisson de paramètre $\varphi(x_p)$. On regroupe les données dans le vecteur $\mathbf{y} \in \mathbb{N}_0^M$ et :

$$p(\mathbf{y} | \lambda, \ell) = \prod_p \exp [y_p \log \varphi(x_p) - \varphi(x_p)] / y_p! , \quad (2)$$

où le produit sur p s'étend sur la liste des pixels observés.

La suite des développements concerne l'inversion dans ce cadre (2) pour des images constantes par classe (1) dont le nombre de classes est connu. Tel quel, le problème est indéterminé, car il possède plus d'inconnues (K amplitudes et P étiquettes) que de données (M observations) et la présence de la distorsion complique encore le problème. La solution décrite repose sur un schéma bayésien incluant les informations disponibles sur les quantités inconnues, comme décrit dans la suite.

2.2 Distributions a priori

L'homogénéité spatiale des étiquettes est prise en compte par un modèle de Potts :

$$p(\ell|\beta) = C(\beta)^{-1} \exp \left[\beta \sum_{p \sim q} \delta(\ell_p; \ell_q) \right] \quad (3)$$

où \sim représente la relation de voisinage entre pixels (ici, classique à quatre voisins). Ce modèle est paramétré par β qui pilote l'extension spatiale des classes.

La distribution fait apparaître le coefficient de normalisation $C(\beta)$. Sa connaissance est cruciale pour estimer β mais son expression analytique compacte est inconnue¹ et son calcul nécessite une gigantesque somme sur les K^P configurations possibles. Cela dit, sur la base de simulations stochastiques, cette fonction a été pré-calculée [4] pour diverses valeurs de K et P . Il s'agit d'un énorme volume de calcul mais qui est effectué une seule fois et sauvegardé.

Concernant le paramètre β , on se place dans un cas où on ne dispose que de peu d'information sur sa valeur et on considère une densité a priori uniforme sur un intervalle $[0, B]$:

$$p(\beta) = \mathcal{U}_{[0, B]}(\beta), \quad (4)$$

où B définit la valeur maximale admissible pour β .

Pour les amplitudes, on s'intéresse également à une situation où peu d'information est disponible. On modélise les paramètres par des variables a priori mutuellement indépendantes $p(\lambda) = \prod_k p(\lambda_k)$. De plus, pour simplifier les développements ultérieurs, on choisit pour chacune une distribution Gamma

$$p(\lambda_k) = \Gamma(a_0)^{-1} b_0^{a_0} \lambda_k^{a_0-1} \exp[-b_0 \lambda_k], \quad (5)$$

de paramètre (a_0, b_0) , connus et réglés pour produire une large dispersion (en pratique, $a_0 = b_0 = 10^{-9}$ pour tous les résultats présentés, ce qui correspond à des lois très peu informatives). En particulier, ce choix assure que les lois (2) et (5) sont conjuguées si la distorsion $\varphi(\cdot)$ est absente. Il toutefois important de noter que pour la majorité des cas étudiés où chaque classe est suffisamment peuplée, l'impact de (5) (et de l'hypothèse de indépendantes a priori) est limité.

2.3 Distribution jointe et décision/estimation

Précédemment, on a défini la distribution (2) pour les données ainsi que celles pour les inconnues : (3) pour les étiquettes,

1. sauf dans le cas du modèle de Ising ($K = 2$) pour lequel $C(\beta)$ est connu depuis longtemps et utilisé dans [4] et plusieurs références incluses.

(4) pour le paramètre de Potts et (5) pour les amplitudes. La figure suivante encode les dépendances et indépendances (conditionnelles) entre ces variables.

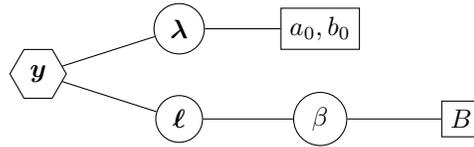


FIG. 1 – Modèle hiérarchique : les cercles et rectangles représentent respectivement les valeurs estimées et fixes.

Il est alors possible de construire la distribution jointe pour l'ensemble des variables :

$$p(\lambda, \ell, \beta, \mathbf{y}) = p(\mathbf{y} | \lambda, \ell) p(\ell|\beta) p(\beta) \prod_k p(\lambda_k). \quad (6)$$

La stratégie bayésienne repose sur le choix d'un coût puis la construction d'un risque comme moyenne du coût sous la loi jointe (données et inconnues) ci-dessus. La fonction de décision (pour les étiquettes) ou d'estimation (pour les amplitudes et le paramètre de Potts) est alors définie comme celle qui minimise le risque [5]. En pratique, on choisit ici un coût binaire pour les erreurs sur les étiquettes et un coût quadratique pour celles sur les paramètres. On obtient ainsi respectivement le maximiseur a posteriori (marginal) et l'espérance a posteriori.

2.4 Échantillonnage de la loi a posteriori

Le calcul de ces quantités repose sur la loi a posteriori, elle même proportionnelle à la jointe (6). Clairement, leur calcul analytique n'est pas possible à cause de la structure emberlificotée de cette loi. On s'appuie alors sur des outils d'échantillonnage stochastique : on génère des échantillons de la loi a posteriori puis on calcule les moyennes et les probabilités empiriques pour approcher les espérances et maxima a posteriori.

Le problème numérique est donc de produire des échantillons sous la loi a posteriori. A nouveau, étant donnée sa complexité, il n'est pas possible de les produire directement et on utilise un algorithme de Gibbs [5, 6]. Il permet de découper le problème global en trois sous problèmes plus simples : on échantillonne successivement chacune des variables λ , ℓ et β , dans un schéma itératif, sous leurs lois a posteriori conditionnelles. Celles-ci sont proportionnelles à la distribution jointe (6) et il suffit donc de retenir seulement les facteurs incluant le paramètre considéré. On sait qu'on produit ainsi une chaîne de Markov dont la loi tend vers la loi d'intérêt. Un pseudo-code est donné ci-après (Algorithme 1). La suite de cette section concerne le développement des échantillonneurs.

2.4.1 Amplitudes des classes λ

La loi des amplitudes se déduit de (6), i.e., $p(\lambda_k | \ell, \mathbf{y}) \propto \exp[-\psi(\lambda_k)]$, où $\psi(\lambda_k)$ représente une énergie définie par

$$\psi(\lambda_k) = \chi(\lambda_k) + b_0 \lambda_k - (a_0 - 1) \log \lambda_k,$$

avec $\chi(\lambda_k) = N_k \varphi(\lambda_k) - S_k \log \varphi(\lambda_k)$ et S_k (resp. N_k) la somme (resp. le nombre) de pixels observés et associés à la classe k . On pose également $m_k = S_k/N_k$.

Sans distorsion, i.e., $\varphi(\lambda) = \lambda$, on trouve une loi Gamma de paramètres $a = a_0 + S_k$ et $b = b_0 + N_k$. Lorsque la distorsion est présente, cette loi n'est pas standard et on ne sait pas l'échantillonner directement. Parmi les algorithmes disponibles [5, 6], nous nous appuyons sur une étape de Metropolis-Hasting : sommairement, on tire une valeur sous une autre loi (dite de proposition) puis, avec probabilité précisément déterminée, on accepte la valeur tirée ou on duplique la valeur précédente. Dans cette famille, plusieurs options sont possibles pour la loi de proposition : indépendante, marche aléatoire, éventuellement directionnelle [7, 8],... Ici, nous explorons la version indépendante et nous approfondissons le choix de la proposition en vue d'améliorer les performances en terme de taux d'acceptation et de mélange. Pour cela, on étudie la fonction χ et on montre qu'elle possède un unique minimiseur $\tilde{\lambda}$ et on détermine la courbure à ce point :

$$\tilde{\lambda} = s \operatorname{ath}(m_k/s) / \alpha \quad \text{et} \quad \chi''(\tilde{\lambda}) = N_k \alpha^2 \left[1 - \left(\frac{m_k}{s} \right)^2 \right] / m_k$$

à partir de quoi, on définit une loi de proposition $q(\lambda_k)$ approchant la loi désirée. Plus précisément, on approche $\exp[-\chi(\lambda_k)]$ par une loi Gamma de moyenne $\tilde{\lambda}$ et précision $\chi''(\tilde{\lambda})$ (Méthode 1) ou de mode $\tilde{\lambda}$ et de courbure $\chi''(\tilde{\lambda})$ (Méthode 2). Cette loi Gamma est multipliée à la loi a priori Gamma pour obtenir une loi de proposition Gamma.

Nous avons également développé une autre approche (Méthode 3) qui repose sur une approximation globale de la loi par minimisation de la divergence de Kullback-Leiber

$$KL(p(\lambda_k | \ell, \mathbf{y}) || q(\lambda_k)), \quad (7)$$

par rapport à $q(\lambda_k)$ qui est définie comme une loi Gamma. La minimisation de (7) s'effectue en satisfaisant $E_{p(\lambda_k | \ell, \mathbf{y})} [\lambda_k] = E_{q(\lambda_k)} [\lambda_k]$ et $E_{p(\lambda_k | \ell, \mathbf{y})} [\ln(\lambda_k)] = E_{q(\lambda_k)} [\ln(\lambda_k)]$, où les espérances par rapport à $p(\lambda_k | \ell, \mathbf{y})$ sont calculées par intégration numérique (intégrales univariées).

2.4.2 Étiquettes ℓ

La loi a posteriori pour l'ensemble des étiquettes présente une structure markovienne de même système de voisinage que

Algorithm 1 — Gibbs incluant Metropolis-Hastings

```

1: Initialisation  $\beta^{(0)}$ 
2: for  $q = 1, 2, \dots$  do
3:   for  $k = 1, 2, \dots, K$  do
4:     Target :  $p(\lambda_k | \ell^{(q-1)}, \mathbf{y})$ 
5:     Sample a proposition  $\lambda_k^p$ 
6:     Compute the acceptance probability
7:     Accept  $(\lambda_k^{(q)} = \lambda_k^p)$  or Duplicate  $(\lambda_k^{(q)} = \lambda_k^{(q-1)})$ 
8:   end for
9:   Sample  $\ell^{(q)}$  under  $p(\ell | \lambda^{(q)}, \beta^{(q-1)}, \mathbf{y})$ 
10:  Sample  $\beta^{(q)}$  under  $p(\beta | \ell^{(q)})$ 
11: end for

```

la distribution a priori. La grille de pixels constitue un damier, définissant deux sous-ensembles, disons ébène et ivoire. Sur cette grille, les pixels ébènes sont indépendants étant donnés les pixels ivoire et réciproquement. On échantillonne donc en parallèle les pixels de chaque sous-ensemble.

Pour un pixel p , la probabilité de la classe $k \in \{1, \dots, K\}$ se déduit aussi de (6). Selon qu'il est observé ou pas :

$$p(\ell_p = k | \star) \propto \begin{cases} \exp[y_p \log \varphi(\lambda_k) - \varphi(\lambda_k)] \exp[\beta N_{p,k}] \\ \exp[\beta N_{p,k}] \end{cases}$$

où \star représente « toutes les autres variables » et $N_{p,k}$ le nombre de voisin du pixel p possédant l'étiquette k . L'échantillonnage de ℓ_p se fait donc en tirant suivant cette loi (normalisée à 1).

2.4.3 Paramètre de Potts β

Pour terminer, la loi conditionnelle a posteriori de β est :

$$p(\beta | \ell) \propto C(\beta)^{-1} \exp \left[\beta \sum_{p \sim q} \delta(\ell_p; \ell_q) \right] \mathcal{U}_{[0, B]}(\beta).$$

On tire avantage du fait qu'il s'agit d'une variable scalaire sur un intervalle fini et que la fonction de partition $C(\beta)$ a été pré-calculée [4] sur une grille² de valeurs de β . Il est alors aisé de calculer la fonction de répartition $F(\beta)$ par intégration numérique et interpolation linéaire. Il suffit alors de tirer un échantillon u uniforme sur $[0, 1]$ et de calculer $\beta = F^{-1}(u)$.

3 Résultats numériques

On considère une image vraie x^* carrée de $P = 128^2$ pixels et $K = 3$ classes. Les vraies amplitudes sont $\lambda^* = [5, 30, 70]$ et la distorsion est $\varphi(u) = s \tanh(\alpha u/s)$ avec $\alpha = 1$ et $s = 50$. Par ailleurs, le taux de pixel manquant est de 1/2. Les images vraies et observées sont présentées sur la Fig. 2. Il est impor-

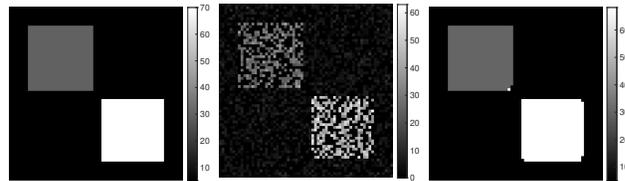


FIG. 2 – De gauche à droite : vraie x^* , données \mathbf{y} , estimée \hat{x}

tant de noter que nous avons également fait fonctionner l'algorithme de nombreuses fois, à partir d'initialisations identiques et différentes ainsi que pour d'autres valeurs de P et K ainsi que de λ^* et x^* . Nous avons systématiquement observé un comportement convergent et cohérent. Le temps de calcul³ est de l'ordre de 3 secondes pour environ 1000 itérations.

Fig. 2 présente les résultats d'estimation de l'image : l'image estimée \hat{x} est visuellement quasiment identique à x^* , en géométrie et en amplitude. Tab. 1 compare les valeurs d'amplitude dont les estimées sont très proches des vraies valeurs. Il

2. En pratique, on fixe $B = 3$ et la grille a un pas de 0.01.

3. sous Matlab, sur un PC standard à 4 GHz CPU et 64 GB de RAM.

n'existe pas de vraie valeur pour β car les étiquettes traitées ne sont pas générées suivant (3) mais on obtient $\hat{\beta} \approx 1.4$.

| Paramètre | λ_1 | λ_2 | λ_3 | β |
|--------------|-------------|-------------|-------------|---------|
| Vraie valeur | 5 | 30 | 70 | — |
| Estimées | 5.05 | 28.40 | 68.60 | 1.41 |
| Écart-type | 0.06 | 0.42 | 2.01 | 0.03 |

TAB. 1 – Paramètres : valeurs vraies et valeurs estimées.

Fig. 3 montre les erreurs de d'étiquetage : leur nombre est insignifiant (moins de 0.1%). Cette figure montre, à droite, les probabilités associées aux étiquettes sélectionnées. La plupart des probabilités est proche de 1, indiquant une décision « sure » et les probabilités plus faibles aux bords des régions indiquent une moins bonne fiabilité. Par ailleurs, les pixels incorrectement étiquetés se situent dans ces régions de faibles probabilités dont l'identification représente un des intérêts majeurs des méthodes bayésiennes.

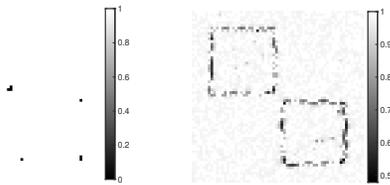


FIG. 3 – Erreur de sélection (en noir) à gauche et probabilité (marginale) a posteriori de l'étiquette sélectionnée à droite.

Approfondissement concernant la proposition

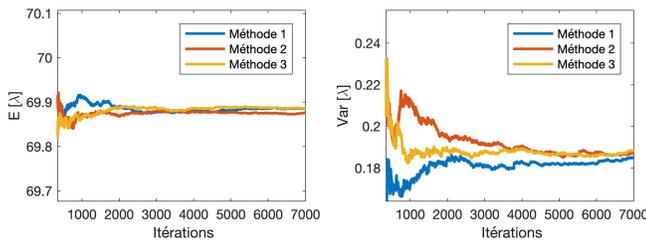


FIG. 4 – Estimation de la moyenne et de la variance d'une loi conditionnelle de λ en utilisant $M = 5000$ pixels générés suivant (2) avec $(\lambda, \alpha, s) = (70, 1, 50)$.

Enfin, on s'intéresse plus particulièrement aux trois méthodes d'échantillonnage des amplitudes discutées dans la Section 2. On considère par exemple ici, un ensemble de $M = 5000$ pixels observés, associés à une même classe (cf. valeurs des paramètres fixes sur la Fig. 4). Cette figure illustre la qualité d'estimation de la moyenne et de la variance de λ en fonction du nombre d'échantillons générés. Les 300 premiers échantillons sont écartés et les estimations sont produites à partir des autres échantillons (au moins 50). Ici, les 3 méthodes proposées donnent généralement des résultats très satisfaisants et similaires, se rapprochant d'un échantillonnage parfait (taux d'acceptation $> 98\%$, bien meilleurs qu'avec une marche aléatoire bien ajustée). Il est cependant à noter que la Méthode 3 a un

coût calculatoire supérieur (calcul numérique d'intégrales). Cependant, une comparaison plus complète (omise ici par manque de place) est nécessaire pour confirmer la robustesse de ces approches dans un contexte plus large (dans l'échantillonneur global, avec des non-linéarités potentiellement non-dérivables).

4 Synthèse et perspectives

Ce travail propose une approche bayésienne pour la segmentation d'images dégradées (non-linéarité, pixels manquants) et contaminées par un bruit Poisson dans un régime faible flux. Il définit des estimateurs optimaux et utilise des méthodes de simulation stochastique pour leurs calculs. Les résultats sont particulièrement attractifs concernant la segmentation et la restitution des amplitudes, malgré les difficultés du problème. Les travaux en cours et futurs incluent d'autres modèles de bruit de comptage [9, 10] et la présence de dégradation spatiale (flou) [4].

Remerciements — Les auteurs remercient le GIE CentraleSupélec pour son financement « travaux champ de Potts » et la Royal Academy of Engineering (Research Fellowship RF 201617/16/31).

Références

- [1] J. C. Russ, *The Image Processing Handbook (Seventh Edition)*, CRC Press, 2015.
- [2] Z. Kato et J. Zerubia, *Markov Random Fields in Image Segmentation*, Now publishers Inc, 2012.
- [3] J.-F. Giovannelli et A. Barbos, « Unsupervised segmentation of piecewise constant images from incomplete, distorted and noisy data », in *Proc. of the Int. Conf. on Stat. Signal Proc.*, Palma de Majorque, Spain, June 2016, pp. 1–5.
- [4] C. Vacar et J.-F. Giovannelli, « Unsupervised joint deconvolution and segmentation method for textured images : a Bayesian approach and an advanced sampling algorithm », *EURASIP Journal on Advances in Signal Processing*, jan 2019.
- [5] C. P. Robert, *The Bayesian Choice. From decision-theoretic foundations to computational implementation*, Springer Texts in Statistics. Springer Verlag, New York, NY, USA, 2007.
- [6] S. Brooks, A. Gelman, G. L. Jones, et X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, Chapman & Hall / CRC, Boca Raton, USA, 2011.
- [7] C. Vacar, J.-F. Giovannelli, et Y. Berthoumieu, « Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance », in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 3964–3967.
- [8] M. Girolami et B. Calderhead, « Riemannian manifold Hamiltonian Monte Carlo (with discussion) », *J. R. Statist. Soc. B*, vol. 73, pp. 123–214, 2011.
- [9] Y. Altmann, S. McLaughlin, et M. Padgett, « Unsupervised restoration of subsampled images constructed from geometric and binomial data », in *Proc. IEEE CAMSAP*, Curacao, Dec 2017, pp. 1–5.
- [10] Y. Altmann, R. Aspden, M. Padgett, et S. McLaughlin, « A bayesian approach to denoising of single-photon binary images », *IEEE Trans. on Computat. Imaging*, vol. 3, no. 3, pp. 460–471, Sep. 2017.