



# PARAMÉTRISATION « TÉLÉGRAPHIQUE » DE CHAMPS DE PICKARD POUR LA SEGMENTATION NON-SUPERVISÉE D'IMAGES

Gildas PERROT<sup>†</sup>, Yves GOUSSARD<sup>†</sup> et Jérôme IDIER<sup>‡</sup>

<sup>†</sup>École polytechnique  
Institut de génie biomédical  
C.P. 6079, Succursale « Centre-ville »  
Montréal (Québec) H3C 3A7, Canada

<sup>‡</sup>Laboratoire des signaux et systèmes  
École supérieure d'électricité  
Plateau de Moulon  
91192 GIF-SUR-YVETTE Cédex, France

## RÉSUMÉ

Cette communication présente une méthode de segmentation non-supervisée d'images modélisées comme des champs de Markov unilatéraux à niveaux discrets. Devijver & Dekesel ont montré que l'emploi de tels champs conduit à des algorithmes de faible complexité numérique, tant pour la segmentation que pour l'apprentissage des hyperparamètres. Notre contribution réside dans l'introduction d'une paramétrisation « télégraphique » du champ de Markov unilatéral. Celle-ci assure que certaines propriétés théoriques du champ (stationnarité en particulier) sont bien vérifiées. Elle permet de plus une réduction sensible du nombre de paramètres à estimer au cours de l'apprentissage, facilitant ainsi la convergence et la précision des méthodes d'estimation employées.

## ABSTRACT

This communication presents a non-supervised segmentation method based upon a discrete-level unilateral Markov field model of the image. Devijver & Dekesel showed that such models yield numerically efficient algorithms, for segmentation and for hyperparameter estimation as well. Our contribution lies in the derivation of a *telegraphic* parameterization of the unilateral Markov field. This parameterization ensures that some theoretical properties of the field (e.g., stationarity) do hold. In addition, it decreases the number of hyperparameters that must be estimated, thereby improving convergence speed and accuracy of the corresponding estimation methods.

## 1 Introduction

Cette communication traite de la segmentation non-supervisée d'images modélisées comme des champs markoviens. Les champs markoviens sont d'un grand intérêt en segmentation car ils modélisent explicitement une caractéristique importante de l'image à segmenter : la présence de zones homogènes séparées par des discontinuités. L'emploi de telles représentations présente cependant des inconvénients tels que la lourdeur des algorithmes correspondants et la difficulté à résoudre le problème *non-supervisé*, c'est à dire l'estimation des paramètres du champ directement à partir des données observées lors d'une étape d'*apprentissage*.

Pour surmonter ces difficultés, nous proposons une méthode qui constitue une extension des travaux de DEVIJVER & DEKESEL [1]. Ces auteurs ont développé une méthode de segmentation non-supervisée basée sur des champs markoviens unilatéraux particuliers : les champs de PICKARD [2]. De tels champs permettent une mise en œuvre rapide de l'étape de segmentation. Cependant, des difficultés subsistent au niveau de la paramétrisation du champ et de l'étape de réestimation.

La contribution de cette communication est de proposer

une nouvelle paramétrisation de champs de PICKARD ainsi que les formules d'apprentissage correspondantes. L'intérêt de cette paramétrisation, décrite de manière plus détaillée dans une autre soumission [3], est double : d'une part, elle assure que certaines propriétés théoriques du champ sont vérifiées ; d'autre part, les algorithmes correspondants présentent une complexité numérique plus faible et une convergence plus sûre que celle de la méthode décrite dans [1]. La méthode proposée présente donc de meilleures caractéristiques, tant au plan théorique que pratique.

## 2 Méthode

### 2.1 Champs de Pickard

Les champs de MARKOV complètement définis par leur potentiel sur l'ensemble des *cliques maximales*, qui, pour les champs du deuxième ordre, sont constituées par les cellules de quatre pixels de la forme  $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ . De tels champs généraux ont permis le développement de nombreuses méthodes de segmentation. Ils présentent cependant plusieurs inconvénients. En particulier, leur échantillonnage requiert l'emploi d'algorithmes lourds tels que l'échantillonneur de



GIBBS. De plus, de tels champs de sont généralement pas stationnaires, et il est difficile, voire impossible de calculer la distribution de groupes de pixels du champ, ainsi que la corrélation entre pixels.

Afin de remédier à ces inconvénients, PICKARD [2] s'est posé la question de l'existence de champs définis par une mesure homogène  $\tau \left( \begin{smallmatrix} A & B \\ C & D \end{smallmatrix} \right)$  sur toutes les cliques, que l'on puisse simuler de manière unilatérale, et qui soient à la fois markoviens du deuxième ordre et stationnaires. Il a établi les deux résultats suivants :

1° Pour qu'un champ unilatéral généré par une mesure  $\tau$  stationnaire sur  $\left( \begin{smallmatrix} A & B \\ C & D \end{smallmatrix} \right)$  soit markovien, il suffit que

$$\tau(B, C | A) = \tau(B | A)\tau(C | A) \quad (1)$$

2° Pour qu'un tel champ vérifiant (1) soit de plus stationnaire, il faut et il suffit que

$$\tau(B, C | D) = \tau(B | D)\tau(C | D) \quad (2)$$

ou

$$\begin{cases} \tau(A, D | C) = \tau(A | C)\tau(D | C) \\ \tau(A, D | B) = \tau(A | B)\tau(D | B) \end{cases} \quad (3)$$

Un champ vérifiant les deux conditions ci-dessus est appelé champ de PICKARD. De tels champs possèdent des propriétés intéressantes ; en particulier, ils permettent le calcul immédiat d'une forme symétriques des potentiels de Gibbs. D'autre part, le calcul des distributions marginales des lignes et des colonnes est simple, et on montre d'ailleurs que celles-ci sont des chaînes de Markov dont les probabilités initiales et de transition sont données respectivement  $\tau(A)$ ,  $\tau(B | A)$  et  $\tau(C | A)$ .

## 2.2 Segmentation non-supervisée

Pour exploiter de telles propriétés, on modélise l'image par un champ de PICKARD à  $N$  états discrets, et on émet l'hypothèse que la distribution conditionnelle de l'image observée  $\mathbf{Y}$  est donnée par :

$$f_{\mathbf{Y} | \mathbf{X}}(\mathbf{Y} | \mathbf{X}) = \prod_{i,j} f(Y_i^j | X_i^j), \quad (4)$$

$$f(Y_i^j | X_i^j = n) = \mathcal{N}(m_n, \sigma_n^2), \quad (5)$$

où  $i$ ,  $j$  et  $n$  désignent respectivement l'indice des lignes, celui des colonnes et l'état de  $X_i^j$ .  $\mathcal{N}(m, \sigma^2)$  est une distribution gaussienne de moyenne  $m$  et de variance  $\sigma^2$ . Pour effectuer la segmentation, on adopte une approche de type *maximum a posteriori (MAP) marginal* qui consiste à maximiser pour chaque pixel la vraisemblance  $P(X_i^j | \mathbf{Y})$ . DEVIJVER & DEKESEL [1] proposent d'approcher cette vraisemblance par  $P(X_i^j | \mathbf{Y}_i, \mathbf{Y}^j)$  où  $\mathbf{Y}_i$  et  $\mathbf{Y}^j$  désignent respectivement la ligne d'indice  $i$  et la colonne d'indice  $j$  de  $\mathbf{Y}$ . On a immédiatement :

$$P(X_i^j | \mathbf{Y}_i, \mathbf{Y}^j) \propto f(\mathbf{Y}_i, \mathbf{Y}^j | X_i^j) P(X_i^j), \quad (6)$$

et la dépendance entre  $\mathbf{Y}$  et  $\mathbf{X}$  donnée par (4) associée aux propriétés d'indépendance conditionnelle entre pixels des champs de PICKARD permet alors d'écrire

$$P(X_i^j | \mathbf{Y}_i, \mathbf{Y}^j) \propto f(\mathbf{Y}_i | X_i^j) f(\mathbf{Y}^j | X_i^j) P(X_i^j). \quad (7)$$

Grâce à la structure de chaîne de MARKOV de  $\mathbf{X}_i$  et  $\mathbf{X}^j$  ainsi qu'à la forme de la dépendance entre  $\mathbf{Y}$  et  $\mathbf{X}$ , il est possible d'évaluer les deux probabilités conditionnelles du membre de droite de (7) à l'aide d'algorithmes « forward-backward » qui ont démontré leur efficacité dans le domaine de la reconnaissance de la parole.

Ces opérations nécessitent de connaître les paramètres du modèle : il s'agit :

- des paramètres  $\theta_{\mathbf{Y} | \mathbf{X}}$  de la distribution conditionnelle  $f_{\mathbf{Y} | \mathbf{X}}$ . D'après (4) et (5),  $\theta_{\mathbf{Y} | \mathbf{X}}$  est composé des  $N$  couples  $(m_n, \sigma_n^2)$  ;
- des paramètres  $\theta_{\mathbf{X}}$  de la distribution *a priori* du champ de PICKARD qui interviennent dans le deuxième membre de (7). Cette dernière expression ne fait intervenir que les chaînes de MARKOV que constituent les lignes et les colonnes, et ne nécessite donc que la connaissance des probabilités initiales et de transition qui sont données immédiatement par  $\tau(A)$ ,  $\tau(B | A)$  et  $\tau(C | A)$ .

En segmentation non-supervisée, ces paramètres doivent être préalablement estimés à partir de l'image observée  $\mathbf{Y}$ . Vu le cadre probabiliste adopté ici, il est naturel d'avoir recours à un estimateur de type *maximum de vraisemblance* qui possède de bonnes propriétés asymptotiques. L'estimée des paramètres est donc définie par

$$\hat{\theta} \triangleq \arg \max_{\theta} f_{\mathbf{Y}}(\mathbf{Y}; \theta), \quad (8)$$

avec  $\theta \triangleq \{\theta_{\mathbf{X}}, \theta_{\mathbf{Y} | \mathbf{X}}\}$ . Malheureusement, le problème de maximisation ci-dessus n'admet pas de solution explicite, et l'on doit avoir recours aux méthodes numériques d'optimisation. Parmi des dernières, l'algorithme EM permet en général d'allier des propriétés de convergence satisfaisantes à une mise en œuvre raisonnable. Il s'agit d'une méthode qui, partant d'une valeur initiale quelconque  $\hat{\theta}^0$ , génère une suite d'estimées  $\hat{\theta}^k$  en itérant les deux étapes suivantes :

$$\text{Expectation (E)} \quad \text{Exprimer } Q(\theta, \hat{\theta}^k; \mathbf{Y}), \quad (9)$$

$$\text{Maximization (M)} \quad \hat{\theta}^{k+1} = \arg \max_{\theta} Q(\theta, \hat{\theta}^k; \mathbf{Y}) \quad (10)$$

avec

$$Q(\theta, \theta'; \mathbf{Y}) \triangleq E[J(\mathbf{X}, \mathbf{Y}; \theta) | \mathbf{Y}; \theta'], \quad (11)$$

$$J(\mathbf{X}, \mathbf{Y}; \theta) \triangleq \ln f_{\mathbf{Y} | \mathbf{X}}(\mathbf{Y} | \mathbf{X}; \theta_{\mathbf{Y} | \mathbf{X}}) + \ln P(\mathbf{X}; \theta_{\mathbf{X}}). \quad (12)$$

On montre que les  $\hat{\theta}^k$  sont de vraisemblance croissante au sens large, et que l'algorithme converge vers un point critique de la vraisemblance. D'un point de vue pratique, l'intérêt de cette technique provient de ce que (11) et (12) font intervenir  $\mathbf{X}$  explicitement, ce qui permet de mettre en œuvre les étapes E et M sans trop de difficulté.

D'après (12), il est clair que l'étape M se décompose en deux problèmes indépendants de maximisation des quantités  $Q_{\mathbf{Y} | \mathbf{X}}$  et  $Q_{\mathbf{X}}$  définies par

$$Q_{\mathbf{Y} | \mathbf{X}} \triangleq E[\ln f_{\mathbf{Y} | \mathbf{X}}(\mathbf{Y} | \mathbf{X}; \theta_{\mathbf{Y} | \mathbf{X}}) | \mathbf{Y}; \theta'], \quad (13)$$

$$Q_{\mathbf{X}} \triangleq E[\ln P(\mathbf{X}; \theta_{\mathbf{X}}) | \mathbf{Y}; \theta'], \quad (14)$$

respectivement par rapport à  $\theta_{\mathbf{Y}|\mathbf{X}}$  et  $\theta_{\mathbf{X}}$ . Pour les mêmes raisons que celles invoquées dans [3], la maximisation de  $Q_{\mathbf{Y}|\mathbf{X}}$  pose peu de difficultés et conduit à une extension bi-dimensionnelle triviale des formules de réestimation de BAUM-WELCH pour les paramètres  $\theta_{\mathbf{Y}|\mathbf{X}}$ . La maximisation de  $Q_{\mathbf{X}}$  est plus difficile. Afin de simplifier les calculs et de tirer parti de la structure de chaîne de MARKOV des lignes et des colonnes de  $\mathbf{X}$ , DEVIJVER & DEKESEL [1] proposent d'effectuer l'approximation

$$Q_{\mathbf{X}} \approx \frac{1}{2} \left\{ \sum_i E[\ln P(\mathbf{X}_i; \theta_{\mathbf{X}}) | \mathbf{Y}_i; \theta'] + \sum_j E[\ln P(\mathbf{X}^j; \theta_{\mathbf{X}}) | \mathbf{Y}^j; \theta'] \right\}, \quad (15)$$

ce qui revient à considérer comme indépendantes et identiquement distribuées les lignes et les colonnes de toutes les données. On observe alors que chaque terme du second membre de (15) est identique à celui obtenu pour une chaîne de Markov cachée à  $N$  états. DEVIJVER & DEKESEL [1] adoptent une paramétrisation classique de ces chaînes par probabilités initiales et probabilités de transition. Ils aboutissent ainsi à une extension bi-dimensionnelle immédiate des formules de réestimation de BAUM-WELCH [1] pour les paramètres  $\theta_{\mathbf{X}}$ .

Une telle manière de procéder présente deux inconvénients majeurs: d'une part, la représentation non-paramétrique choisie ne garantit pas la stationnarité et la réversibilité de chaque ligne et colonne de  $\mathbf{X}$ , et la mesure  $\tau$  correspondante ne satisfait généralement pas les conditions nécessaires pour générer un champ de PICKARD; d'autre part, le nombre de paramètres à déterminer est de l'ordre de  $N^2$ , ce qui alourdit sensiblement les calculs et provoque des difficultés de convergence de l'algorithme lorsque le nombre  $N$  d'états augmente.

### 2.3 Utilisation d'une paramétrisation « télégraphique »

Pour remédier aux difficultés mentionnées ci-dessus, nous proposons d'effectuer l'approximation (15), mais d'utiliser un modèle « télégraphique » pour les chaînes de MARKOV cachées que constituent des lignes et des colonnes de  $\mathbf{Y}$ . Une chaîne télégraphique paramétrisée par deux vecteurs  $\lambda$  et  $\mu$  de taille  $N$  tels que

$$\mathbf{p} = (\mathbf{I} - \mathbf{\Lambda} + \mu \lambda^T)^{-1} \mu, \quad (16)$$

$$\mathbf{P} = \mathbf{\Lambda} + (\mathbf{1} - \lambda) \mu^T, \quad (17)$$

$$\mathbf{\Lambda} \triangleq \text{diag}(\lambda), \quad (18)$$

$$\mathbf{1} \triangleq (1, \dots, 1)^T, \quad (19)$$

où  $\mathbf{p}$  et  $\mathbf{P}$  désignent respectivement les probabilités initiales et les probabilités de transition de la chaîne.  $\lambda$  et  $\mu$  doivent vérifier les contraintes

$$\mathbf{1}^T \mu = 1 \quad \text{et} \quad \forall n, \mu_n > 0, \quad (20)$$

$$\forall n, -\mu_n / (1 - \mu_n) < \lambda_n < 1. \quad (21)$$

Les chaînes télégraphiques possèdent deux propriétés qui méritent d'être soulignées. D'une part, elles sont nécessairement stationnaires et réversibles;  $\mathbf{p}$  représente donc également la probabilité marginale de chaque élément de la

chaîne. D'autre part, la paramétrisation comporte  $2N - 1$  degrés de liberté, contre  $O(N^2)$  pour une paramétrisation classique. Des détails supplémentaires sur les chaînes télégraphiques peuvent être trouvés dans [3].

Dans notre application, un modèle télégraphique des lignes et des colonnes de  $\mathbf{X}$  présente donc un double avantage:

1° il assure la stationnarité des lignes et des colonnes, qui est l'une des propriétés importantes des champs de PICKARD.  $\mathbf{p}$  représente alors  $\tau(A)$  et  $\mathbf{P}$  représente  $\tau(B | A)$  ou  $\tau(B | A)$  selon que l'on s'intéresse aux colonnes ou aux lignes:

2° il réduit sensiblement le nombre de paramètres du modèle, surtout lorsque  $N$  dépasse 5 ou 6. En segmentation non supervisée, les conditions d'estimation de ces paramètres sont donc plus favorables, et l'on peut espérer une meilleure convergence algorithmes d'optimisation employés.

Tout ceci suppose que l'on puisse évaluer, puis maximiser la version approchée de  $Q_{\mathbf{X}}$  donnée par (15), où  $\theta_{\mathbf{X}}$  est maintenant formé des vecteurs  $\lambda$  et  $\mu$  correspondant aux lignes et aux colonnes. Dans [3], il est montré que, pour une chaîne télégraphique,  $Q_{\mathbf{X}}$  s'exprime à l'aide de  $\lambda$ ,  $\mu$  et des quantités  $\alpha'_n$ ,  $\beta'_n$  et  $s'_n$  définies par

$$\alpha'_n \triangleq \sum_{t=1}^T P(X_t = n; \theta'), \quad (22)$$

$$\beta'_n \triangleq \sum_{t=2}^{T-1} P(X_t = n; \theta'), \quad (23)$$

$$s'_n \triangleq \sum_{t=2}^T P(X_t = n, X_{t-1} = n; \theta'), \quad (24)$$

où  $t$  et  $T$  désignent respectivement l'indice de l'élément courant de la chaîne et le nombre total d'éléments. Il est également montré dans [3] que, au prix d'une approximation mineure, il est possible de maximiser  $Q_{\mathbf{X}}$  par une procédure numérique simple à mettre en œuvre.

Ce résultat peut être étendu au cas de plusieurs chaînes indépendantes identiquement distribuées. Il suffit alors de remplacer dans les définitions (22)-(24) la sommation sur  $t$  par une double sommation sur  $t$  et sur l'indice des chaînes, puis d'appliquer l'algorithme décrit dans [3].

La méthode proposée comporte donc deux étapes. Dans la première, les paramètres du modèle télégraphique sont déterminés par les formules de réestimation présentées dans [3]. Par construction, le modèle est compatible avec les propriétés de la mesure  $\tau$ . La seconde étape permet la segmentation proprement dite de l'image  $\mathbf{Y}$  en utilisant les paramètres estimés à l'étape précédente. La méthode utilisée (MAP marginal) est identique à celle décrite dans [1]. Notons que les deux étapes sont basées sur le même algorithme « forward-backward », ce qui simplifie la mise en œuvre de la méthode.

## 3 Résultats

La méthode proposée a été testée à l'aide de données simulées et réelles. Sur données simulées, nous avons observé un comportement très satisfaisant de la méthode, avec une



convergence rapide de la procédure de réestimation et une bonne estimation des paramètres du modèle. Sur données réelles, la procédure diverge occasionnellement. Il s'agit là d'une conséquence de la dégénérescence de la vraisemblance qui se produit quand on estime conjointement les moyennes et les variances d'un mélange [4]. Ici, le problème a été contourné en fixant les moyennes  $m_k$  à des valeurs déterminées à l'avance par des moyennes locales sur  $Y$ . Ainsi modifiée, la procédure produit des résultats de bonne qualité et est capable de segmenter des structures fines, comme illustré à la figure 1.

La comparaison avec la méthode décrite dans [1] confirme que, comme indiqué à la section 2, la méthode que nous proposons présente une convergence plus sûre et plus rapide. Celle-ci se présente donc comme une alternative intéressante aux techniques existantes. Nos efforts portent maintenant sur la réduction des approximations faites lors de l'évaluation de la vraisemblance marginale.

## Références

- [1] P. DEVIJVER et M. DEKESEL, "Champs aléatoires de Pickard et modélisation d'images digitales," *Traitement du Signal*, vol. 5, pp. 131-150, 1988.
- [2] D. PICKARD, "Unilateral Markov fields," *Adv. Appl. Prob.*, vol. 12, pp. 655-671, 1980.
- [3] J. IDIER et Y. GOUSSARD, "Formules de réestimation pour un modèle de chaîne de Markov cachée stationnaire réversible," in *Proc. GRETSI Conf.*, (Juan-les-Pins, France), 1995.
- [4] A. NÁDAS, "Hidden Markov chains, the forward-backward algorithm, and initial statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 504-506, 1983.

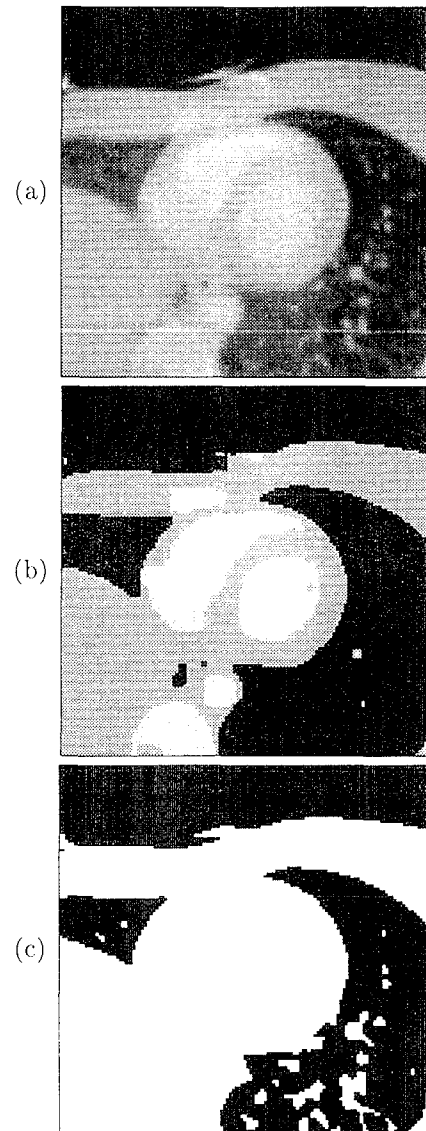


FIG. 1 - Exemple de résultat obtenu sur une image de tomographie axiale. L'objectif est de segmenter l'intérieur des ventricules cardiaques. En dépit du faible contraste de l'image observée (a), la méthode que nous proposons (b) fournit des résultats adéquats alors que la méthode EM décrite dans [1] (c) diverge.