UNSUPERVISED IMAGE SEGMENTATION USING A TELEGRAPH PARAMETERIZATION OF PICKARD RANDOM FIELDS

Yves Goussard

École Polytechnique
Biomedical Engineering Institute
C.P. 6079, Station Centre-Ville
Montréal (Québec)
H3C 3A7, Canada

Jérôme Idier and Alain DeCesare

Laboratoire des Signaux et Systèmes
École Supérieure d'Électricité
Plateau de Moulon
91192 Gif-sur-Yvette Cedex
France

ABSTRACT

This communication presents a non-supervised segmentation method based upon a discrete-level unilateral Markov field model of the image. Such models have been shown to yield numerically efficient algorithms, for segmentation and for hyperparameter estimation as well. Our contribution lies in the derivation of a parsimonious *telegraphic* parameterization of the unilateral Markov field. On a theoretical level, this parameterization ensures that some important properties of the field (e.g., stationarity) do hold. On a practical level, it reduces the computational complexity of the algorithm used in the segmentation and parameter estimation stages of the precedure. In addition, it decreases the number of hyperparameters that must be estimated, thereby improving convergence speed and accuracy of the corresponding estimation method.

1. INTRODUCTION

This communication deals with segmentation of images modeled as Markov random fields (MRFs). MRFs have proved useful in image segmentation because they can explicitly model important features of actual images, such as the presence of homogeneous regions separated by sharp discontinuities. However, the corresponding methods are often computationally intensive and solving the *unsupervised* problem, i.e., estimation of the MRF parameters, generally presents great difficulties.

In order to overcome such difficulties, Devijver and Dekesel [1] proposed an unsupervised segmentation approach in which the image model belongs to a special class of *unilateral* MRFs: Pickard random fields (PRFs). Such models result in a significant reduction of the computational burden, particularly in the segmentation stage. However, several difficulties remain: on a theoretical level, tractable expressions of parametric constraints have yet to be established for n-ary PRFs; in the estimation stage, theoretical properties of PRFs such as stationarity are not enforced; on a more practical level, convergence of the estimation procedure is somewhat questionable and often slow.

In this communication, we propose an extension of Devijver and Dekesel's technique based upon a parsimonious *telegraph* parameterization of PRFs which guarantees that desirable theoretical properties of PRFs are fulfilled. Formulas for model parameter estimation as well as for image segmentation are derived. The parsimonious character of the parameterization results in a significant reduction of the amount of computations and in a better robustness of the estimation procedure. Consequently, the proposed method presents better characteristics from both theoretical and practical standpoints.

2. APPROACH

Our approach is akin to that presented in [1]. The image to be segmented is modeled as a n-ary PRF $\boldsymbol{X}$. Such models are stationary and their joint probability is determined by a measure τ on a four-pixel elementary cell $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ that must fulfill several symmetry and independence conditions [2]. As a consequence, the marginal probability of each row and column of $\boldsymbol{X}$ presents the structure of a stationary and reversible Markov chain whose initial and transition probabilities can easily be deduced from τ.

It is assumed that the observed image $\boldsymbol{Y}$ is a noise-corrupted version of a n-ary PRF $\boldsymbol{X}$, and that the conditional probability distribution of pixel Y_i^j is given by

$$p(\boldsymbol{Y}|\boldsymbol{X}) = \prod_{i,j} p(Y_i^j|X_i^j)\,, \qquad (1)$$

$$p(Y_i^j|X_i^j = k) = f_k\,, \qquad (2)$$

where i, j and k respectively denote the row, column and state indices of the current pixel. Segmentation is performed using a marginal maximum *a posteriori* approach which consists of maximizing marginal likelihood $p(X_i^j|\boldsymbol{Y})$ for each pixel X_i^j. The key to the derivation of a numerically efficient algorithm lies in the following approximation:

$$p(X_i^j|\boldsymbol{Y}) \approx p(X_i^j|\boldsymbol{Y}_i, \boldsymbol{Y}^j)\,, \tag{3}$$

where $\boldsymbol{Y}_i$ and $\boldsymbol{Y}^j$ respectively denote the i-th row and j-th column of $\boldsymbol{Y}$. Applying the Baye's rule and using the orthogonality properties of measure τ immediately yield

$$p(X_i^j|\boldsymbol{Y}_i, \boldsymbol{Y}^j) \propto p(\boldsymbol{Y}_i|X_i^j)p(\boldsymbol{Y}^j|X_i^j)p(X_i^j). \tag{4}$$

Taking advantage of the Markov chain structure of $\boldsymbol{X}_i$ and $\boldsymbol{X}^j$, the first two terms of the right hand side of (4) can be evaluated in an efficient manner by means of forward-backward algorithms [1]. It should be stressed that the type of parameterization of Markov chains $\boldsymbol{X}_i$ and $\boldsymbol{X}^j$ has a significant impact on the computational cost of the forward-backward procedures.

The estimation stage, i.e., the determination of parameter vector Θ which controls the probability distributions f_k and the Markov chain measures, is carried out using a maximum likelihood (ML) approach. The maximum of likelihood function $p(\boldsymbol{Y}|\Theta)$ cannot be expressed in closed form. However, because of the Markov chain structure of the rows and columns of $\boldsymbol{X}$, an expectation-maximization (EM) approach is well suited to iterative maximization of the likelihood function [3]. Starting from an initial value $\hat{\Theta}^0$, the EM algorithm generates a series of estimates $\hat{\Theta}^k$ by iterating the following two steps:

$$\text{E}: \quad \text{Evaluate } Q(\Theta, \hat{\Theta}^k; \boldsymbol{Y})\,, \tag{5}$$

$$\text{M}: \quad \hat{\Theta}^{k+1} = \arg\max_{\Theta} Q(\Theta, \hat{\Theta}^k; \boldsymbol{Y})\,, \tag{6}$$

with

$$Q(\Theta, \Theta'; \boldsymbol{Y}) \triangleq E[J(\boldsymbol{X}, \boldsymbol{Y}; \Theta)|\boldsymbol{Y}; \Theta']\,, \tag{7}$$

$$J(\boldsymbol{X}, \boldsymbol{Y}; \Theta) \triangleq \ln p(\boldsymbol{Y}|\boldsymbol{X}; \Theta) + \ln p(\boldsymbol{X}; \Theta)\,. \tag{8}$$

The procedure can be shown to increase the likelihood until a critical point of the likelihood function is reached. It should be underlined that, in [1], the distributions of $\boldsymbol{X}_i$ and $\boldsymbol{X}^j$ are parameterized in a standard manner by the initial and transition probabilities. Consequently, stationarity and reversibility of each row and column is not guaranteed, and measure τ generally does not fulfill the Pickard conditions. In addition, $O(n^2)$ parameters must be estimated, which significantly increases the computational burden and induces convergence difficulties when the number of states increases (see Section 4).

In order to alleviate these theoretical and practical difficulties, we propose a *telegraph model* (TM) for the rows and columns of $\boldsymbol{X}$, and we derive the corresponding reestimation formulas.

3. TELEGRAPH MODEL AND REESTIMATION FORMULAS

The TM adopted here is a straightforward generalization of a class of of Markov chains proposed in [4] for segmentation of seismic signals. The initial probability vector $\boldsymbol{p}$ and transition probability matrix $\boldsymbol{P}$ of each row and column is parameterized with two vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ such that

$$\boldsymbol{P} = \boldsymbol{\Lambda} + (\boldsymbol{1} - \boldsymbol{\lambda})\boldsymbol{\mu}^T\,, \qquad \boldsymbol{p} = \left(\boldsymbol{I} - \boldsymbol{\Lambda} + \boldsymbol{\mu}\boldsymbol{\lambda}^T\right)^{-1}\boldsymbol{\mu}\,, \tag{9}$$

with $\boldsymbol{\lambda} \triangleq \text{vect}(\lambda_k)$, $\boldsymbol{\Lambda} \triangleq \text{diag}(\lambda_k)$, $\boldsymbol{1} = (1, \ldots, 1)^T$ and $\boldsymbol{I}$ = identity matrix. It can be verified that matrix $\text{diag}(\boldsymbol{p})\boldsymbol{P}$ is symmetric, and that the conditions for $\boldsymbol{p}$ and $\boldsymbol{P}$ to be a valid probability vector and a valid probability transition matrix are given by

$$\forall k,\ \mu_k \geq 0\,;\ \sum_{k=1}^{n} \mu_k = 1\,;\ \forall k,\ -\frac{\mu_k}{1-\mu_k} \leq \lambda_k \leq 1\,. \tag{10}$$

Therefore, as long as the initial state probability vector is equal to $\boldsymbol{p}$ and that constraints (10) are fulfilled, (9) defines a stationary and reversible Markov chain that we choose to parameterize with $\theta = \{\boldsymbol{\lambda}, \boldsymbol{\mu}\}$.

The segmentation stage is carried out in the same way as in [1]. The interest of the TM lies in a simplification of the forward-backward algorithm used to evaluate the approximate marginal likelihood values $p(X_i^j|\boldsymbol{Y}_i, \boldsymbol{Y}^j)$. Each recursion of the algorithm requires the computation of conditional probabilities of the form $p(X_i^j|X_{i-1}^j)$ or $p(X_i^j|X_i^{j-1})$, which directly depend on $\boldsymbol{P}$. Expressing $\boldsymbol{P}$ as a function of vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ allows us to bring the computational complexity down from $O(n^2)$ to $O(2n)$.

Maximization of the exact likelihood $p(\boldsymbol{Y}|\Theta)$ does not yield any tractable expression of the estimates of the model parameters. Following Devijver and Dekesel [1], we approximate the exact likelihood by the pseudo-likelihood $\tilde{p}(\boldsymbol{Y}|\Theta)$ defined by:

$$\tilde{p}(\boldsymbol{Y}|\Theta) \triangleq \prod_i p(\boldsymbol{Y}_i|\Theta) \prod_j p(\boldsymbol{Y}^j|\Theta).$$

Then, an EM algorithm for iterative maximization of $\tilde{p}$ can be derived. The major difficulty lies in the M step which, according to (5)-(8), consist of maximizing quantity $\mathcal{Q}$ defined by

$$\mathcal{Q}(\Theta,\Theta') \stackrel{\triangle}{=} \sum_{\boldsymbol{X}_{(\cdot)}} \ln p(\boldsymbol{X}_{(\cdot)},\boldsymbol{Y}_{(\cdot)};\Theta)\, p(\boldsymbol{X}_{(\cdot)} \mid \boldsymbol{Y}_{(\cdot)};\Theta')\,,$$

where $\boldsymbol{X}_{(\cdot)}$ and $\boldsymbol{Y}_{(\cdot)}$ respectively denote any row or column of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Under our assumptions, maximization of $\mathcal{Q}$ yields two independent optimization problems. The first one applies to the parameters of $\{f_k\}$ and as long as weak orthodoxy conditions hold, one actually obtains the classical Baum-Welch reestimation formulas [5]. The second one applies to TM parameters θ whose new value $\overline{\theta}$ is obtained through maximization of $Q(\theta,\Theta') = \sum_{\boldsymbol{X}_{(\cdot)}} \ln p(\boldsymbol{X}_{(\cdot)};\theta)\, p(\boldsymbol{X}_{(\cdot)} \mid \boldsymbol{Y}_{(\cdot)};\Theta')$. Using the definition of a TM, Q can be expressed as

$$Q(\theta,\Theta') \propto \sum_{k=1}^{n} \Bigg(\alpha_k \ln \mu_k + \beta_k \ln(1-\lambda_k) + s_k \ln \frac{\lambda_k + \mu_k - \lambda_k\mu_k}{\mu_k - \lambda_k\mu_k} \Bigg) - \ln \left(\sum_{k=1}^{n} \frac{\mu_k}{1-\lambda_k} \right), \quad (11)$$

where the general form of quantities α_k, β_k and s_k is given by:

$$\alpha_k \stackrel{\triangle}{=} \sum_{j=1}^{J} p(X_{(\cdot)}^{j} = k | \boldsymbol{Y}_{(\cdot)};\Theta')\,, \quad (12)$$

$$\beta_k \stackrel{\triangle}{=} \sum_{j=2}^{J-1} p(X_{(\cdot)}^{j} = k | \boldsymbol{Y}_{(\cdot)};\Theta')\,, \quad (13)$$

$$s_k \stackrel{\triangle}{=} \sum_{j=2}^{J} p(X_{(\cdot)}^{j-1} = k, X_{(\cdot)}^{j} = k | \boldsymbol{Y}_{(\cdot)};\Theta')\,. \quad (14)$$

These quantities can be evaluated with the same forward-backward algorithm as the one used in the segmentation stage. Maximization of Q with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ under constraints (10) does not yield any tractable expression. However, the last term of the right-hand-side of (11) becomes small with respect to the first term as the size of the rows or columns increases. Neglecting thIS last term and maintaining the expression of Q invariant by index-reversion allows us to make the approximation

$$Q(\theta,\Theta') \approx \tilde{Q}(\theta,\Theta') = \sum_{k=1}^{n} Q_k\,,$$

with

$$Q_k = \eta_k \ln \mu_k(1-\lambda_k) + s_k \ln \left(1 + \frac{\lambda_k}{\mu_k(1-\lambda_k)} \right), \quad (15)$$

where $\eta_k \stackrel{\triangle}{=} (\alpha_k + \beta_k)/2$. Maximization of $\tilde{Q}(\theta,\Theta')$ under equality constraint $\sum_{k=1}^{n} \mu_k = 1$ involves several stages. We first notice that, for a given value of $\boldsymbol{\mu}$, the optimal value of $\boldsymbol{\lambda}$ is obtained through independent maximization of each Q_k. This yields

$$\overline{\lambda}_k = \frac{s_k/\eta_k - \mu_k}{1-\mu_k}\,, \quad (16)$$

and $\overline{\boldsymbol{\lambda}}$ necessarily fulfills the third constraint of (10). Substituting (16) into (15) allows us to express Q_k as a function of μ_k to within an additive constant factor:

$$Q_k = \gamma_k \ln \frac{\mu_k}{1-\mu_k}, \text{ with } \gamma_k \stackrel{\triangle}{=} \eta_k - s_k \geq 0\,. \quad (17)$$

The Lagrange multiplier technique is used for maximization of $\tilde{Q}$ with respect to $\boldsymbol{\mu}$ under constraints (10). Equating the gradient of the corresponding criterion to zero yields:

$$\forall k \in [1,n],\ \ \nu\overline{\mu}_k{}^2 - \nu\overline{\mu}_k + \gamma_k = 0\,, \quad (18)$$

where ν denotes the Lagrange multiplier. When $\nu > 4\gamma_k$, the above equation has two distinct roots located on either side of $1/2$ and respectively denoted by $\mu_k^+(\nu)$ and $\mu_k^-(\nu)$. Thus, 2^n combinations should be compared. However, in order for (10) to be fulfilled, $\overline{\boldsymbol{\mu}}(\nu)$ may only contain one $\mu_k^+(\nu)$. This brings the number of possible combinations down to $n+1$. Furthermore, detailed investigation of the properties of combinations that include one $\mu_k^+(\nu)$ reveals that $\tilde{Q}$ is maximized by the combination that contains $\mu_{\overline{k}}^+(\nu)$ where state $\overline{k}$ is defined by

$$\overline{k} \stackrel{\triangle}{=} \arg\max_k \gamma_k\,. \quad (19)$$

Only two possible combinations remain: the one made up of all $\mu_k^-(\nu)$ and the one containing $\mu_{\overline{k}}^+(\nu)$. Further analysis of their properties shows that one and only one combination fulfills constraints (10) and that it can be selected according to a very simple inequality criterion [6]. Finally, the value of the Lagrange multiplier must be determined. It cannot be expressed in closed form, but tight lower and upper bounds can be easily derived. Classical numerical interpolation techniques can then be employed so as to fulfill constraint $\sum_{k=1}^{n} \overline{\mu}_k = 1$ with arbitrary precision.

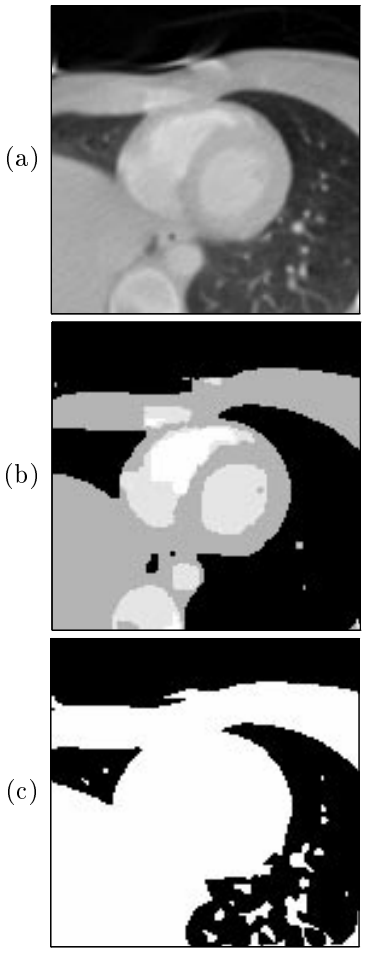

Figure 1: Example of result obtained with a X-ray tomography image. The goal was to segment the inner part of the left and right ventricles. In spite of the low contrast of the observed image (a), the proposed method provides satisfactory results (b) whereas the technique described in [1] (c) diverges.

In the resulting unsupervised segmentation procedure, the first consists of estimating the model parameters using the above EM algorithm. By construction, the model is consistent with the Pickard properties of measure τ. Then, in a second stage, the image is segmented using the parameter values estimated previously. Both stages make use of the same forward-backward algorithm, whose computational complexity is reduced because of the parsimony of the TM parameterization.

4. RESULTS

The proposed method was tested with simulated and real images, under the assumption that the noise distribution (probabilities f_k) is Gaussian. With simulated data, we observed a satisfactory behavior of the method, with fast convergence and accurate results of the estimation procedure. With real data, occasional divergence of the estimation procedure was observed. This may be interpreted as a consequence of a degenerate likelihood function which may occur when the mean and variance of a probability distribution are jointly estimated [7]. Here, the problem was overcome by pre-estimating the mean values m_k using local averages on $\boldsymbol{Y}$. In this manner, results were satisfactory ant he method was able to segment small structures, as illustrated in Figure 1

Comparison with the procedure described in [1] confirms that the method proposed here presents a smaller computational complexity and a faster and more robust convergence. It therefore appears as an interesting alternative to existing unsupervised segmentation methods.

5. REFERENCES

[1] P. Devijver and 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] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, pp. 195–239, 1984.

[4] R. Godfrey, F. Muir, and F. Rocca, "Modeling seismic impedance with Markov chains," *Geophysics*, vol. 45, pp. 1351–1372, 1980.

[5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," *Annals of Mathematical Statistics*, vol. 41, pp. 164–171, 1970.

[6] J. Idier and Y. Goussard, "Formules de réestimation pour un modèle de chaîne de Markov cachée stationnaire réversible," in *Proc. GRETSI Conf.*, (Juan-les-Pins, France), pp. 177–180, 1995.

[7] 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.