UNSUPERVISED ADAPTIVE IMAGE SEGMENTATION

Zoltan Kato, Josiane Zerubia, Marc Berthod

INRIA - 2004 Route des Lucioles - BP 93
06902 Sophia Antipolis Cedex - FRANCE
Tel (33) 93 65 78 57 - Fax (33) 93 65 76 43

Wojciech Pieczynski

INT - 9, rue Charles Fourier
91011 Evry Cedex - FRANCE
Tel (33-1) 60 76 40 40 - Fax (33-1) 60 77 65 29

ABSTRACT

This paper deals with the problem of unsupervised Bayesian segmentation of images modeled by Markov Random Fields (MRF). If the model parameters are known then we have various methods to solve the segmentation problem (Simulated Annealing, ICM, etc...). However, when they are not known, the problem becomes more difficult. One has to estimate the hidden label field parameters from the available image only. Our approach consists of a recent iterative method of estimation, called Iterative Conditional Estimation (ICE), applied to a monogrid Markovian image segmentation model. The method has been tested on synthetic and real satellite images.

1. INTRODUCTION

In real life applications, the model parameters are usually unknown, one has to estimate them only from the observable image. From a statistical viewpoint, this means that we want to estimate parameters from random variables whose joint distribution is a mixture of distributions. If we have a realization of the label field then the problem is relatively easy, we have many standard methods to do parameter estimation (Maximum Likelihood, Coding method [1], etc...). Unfortunately, such a realization is not known, so the direct use of such estimation algorithms is impossible. We have to approximate it by some function of the image data, which is the only observable attribute. Some nowadays used algorithms are iterative [2] subsequently generating a labeling, estimating parameters from it, then generating a new labeling using these parameters, etc ...

Herein, we will present a parameter estimation method applied to monogrid MRF models. The proposed algorithm has been tested on image segmentation problems. Comparative tests have been done on noisy synthetic and real satellite images.

2. PARAMETER ESTIMATION

Hereafter, we consider the monogrid MRF segmentation model originally presented in [3] but with unknown parameters. Let us first review the model. We are given the grey-levels $\mathcal{F}$ of an image $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$, which is the only observable attribute. Moreover, we are given a set of labels denoted by $\Lambda = \{0, 1, \ldots, L-1\}$. The problem is to estimate the model parameters Θ and find the MAP estimate of the label field $\mathcal{X}$ among the possible discrete labelings $\Omega = \Lambda^N = \{\omega = (\omega_{s_1}, \ldots, \omega_{s_N}) : \omega_s \in \Lambda\}$. In the case of unknown parameters, the maximization problem becomes

$$(\widehat{\omega}, \widehat{\Theta}) = \arg\max_{\omega, \Theta} P(\omega, \mathcal{F} \mid \Theta). \quad (1)$$

Since this maximization is not tractable, we use the following equations instead [4, 5]:

$$\widehat{\omega} = \arg\max_{\omega} P(\omega, \mathcal{F} \mid \widehat{\Theta}) \quad (2)$$

$$\widehat{\Theta} = \arg\max_{\Theta} P(\widehat{\omega}, \mathcal{F} \mid \Theta) \quad (3)$$

Herein, we are interested in the solution of the ML estimation using Equation (3):

$$\widehat{\Theta} = \arg\max_{\Theta} P(\widehat{\omega}, \mathcal{F} \mid \Theta) \quad (4)$$

The probability at the right hand side can be written as

$$P(\widehat{\omega}, \mathcal{F} \mid \Theta) = P(\mathcal{F} \mid \widehat{\omega}, \Theta) P(\widehat{\omega} \mid \Theta) \quad (5)$$

Using the model defined in [3], the first term is a product of independent Gaussian densities and the second term is a first order MRF, also known as the Potts model in statistical mechanics [6]:

$$P(\widehat{\omega}, \mathcal{F} \mid \Theta) = \prod_{s \in \mathcal{S}} \frac{1}{\sqrt{2\pi}\sigma_{\widehat{\omega}_s}} \exp\left(-\frac{(f_s - \mu_{\widehat{\omega}_s})^2}{2\sigma^2_{\widehat{\omega}_s}}\right) \cdot \frac{\exp(-\beta \sum_{\{s,r\} \in \mathcal{C}} \delta(\widehat{\omega}_s, \widehat{\omega}_r))}{Z(\beta)} \quad (6)$$

$$\text{with } Z(\beta) = \sum_{\omega \in \Omega} \exp\left(-\beta \sum_{\{s,r\} \in \mathcal{C}} \delta(\omega_s, \omega_r)\right) \quad (7)$$

$$\text{and } \delta(\widehat{\omega}_s, \widehat{\omega}_r) = \begin{cases} 0 & \text{if } \widehat{\omega}_s = \widehat{\omega}_r \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

We have $2L + 1$ parameters (two for each class and one hyperparameter β) denoted by the vector Θ. The first $2L$ parameters are estimated from the Gaussian term and the last one is computed from the Markovian term. Instead of

THIS WORK HAS BEEN PARTIALLY FUNDED BY CNES, AFIRST AND DRED/GDR-TDSI.

0-7803-2431-5/95 $4.00 © 1995 IEEE

the likelihood function defined in Equation (6), we consider the simpler logarithmic likelihood:

$$\begin{aligned}\ln(L(\Theta)) &= \sum_{s\in\mathcal{S}}\left(-\ln(\sqrt{2\pi}\sigma_{\widehat{\omega}_s}) - \frac{(f_s-\mu_{\widehat{\omega}_s})^2}{2\sigma^2_{\widehat{\omega}_s}}\right)\\ &\quad -\beta\sum_{\{s,r\}\in\mathcal{C}}\delta(\widehat{\omega}_s,\widehat{\omega}_r) - \ln(Z(\beta)) \qquad (9)\\ &= \underbrace{\sum_{\lambda\in\Lambda}\sum_{s\in\mathcal{S}_\lambda}\left(-\ln(\sqrt{2\pi}\sigma_\lambda) - \frac{(f_s-\mu_\lambda)^2}{2\sigma^2_\lambda}\right)}_{\mathcal{G}(\mu_\lambda,\sigma_\lambda)}\\ &\quad \underbrace{-\beta\sum_{\{s,r\}\in\mathcal{C}}\delta(\widehat{\omega}_s,\widehat{\omega}_r) - \ln(Z(\beta))}_{\mathcal{M}(\beta)} \qquad (10)\end{aligned}$$

where $\mathcal{S}_\lambda$ is the set of pixels where $\widehat{\omega} = \lambda$. To get the minimum of the likelihood function, $\widehat{\Theta}$ must satisfy the following equations:

$$\forall\lambda\in\Lambda: \quad \frac{\partial\mathcal{G}(\mu_\lambda,\sigma_\lambda)}{\partial\mu_\lambda} = 0 \qquad (11)$$

$$\frac{\partial\mathcal{G}(\mu_\lambda,\sigma_\lambda)}{\partial\sigma_\lambda} = 0 \qquad (12)$$

$$\text{and } \frac{\partial\mathcal{M}(\beta)}{\partial\beta} = 0 \qquad (13)$$

The solution with respect to (w.r.t.) μ_λ and σ_λ of the above system is simply the empirical mean and variance:

$$\begin{aligned}\forall\lambda\in\Lambda: \quad \mu_\lambda &= \frac{1}{\mid\mathcal{S}_\lambda\mid}\sum_{s\in\mathcal{S}_\lambda} f_s,\\ \sigma^2_\lambda &= \frac{1}{\mid\mathcal{S}_\lambda\mid}\sum_{s\in\mathcal{S}_\lambda}(f_s-\mu_\lambda)^2. \qquad (14)\end{aligned}$$

The solution w.r.t. β, however, is not as easy. Let us consider the derivative of $\mathcal{M}(\beta)$:

$$\begin{aligned}&\frac{\partial}{\partial\beta}\left(-\beta N^{ih}(\widehat{\omega}) - \ln\left(\sum_{\omega\in\Omega}\exp(-\beta N^{ih}(\omega))\right)\right)\\ &= -N^{ih}(\widehat{\omega}) + \frac{\sum_{\omega\in\Omega}N^{ih}(\omega)\exp(-\beta N^{ih}(\omega))}{\sum_{\omega\in\Omega}\exp(-\beta N^{ih}(\omega))} = 0 \qquad (15)\end{aligned}$$

with $N^{ih}(\widehat{\omega}) = \sum_{\{s,r\}\in\mathcal{C}}\delta(\widehat{\omega}_s,\widehat{\omega}_r)$ is the number of inhomogeneous cliques in $\widehat{\omega}$. From Equation (15), we get:

$$N^{ih}(\widehat{\omega}) = \frac{\sum_{\omega\in\Omega}N^{ih}(\omega)\exp(-\beta N^{ih}(\omega))}{\sum_{\omega\in\Omega}\exp(-\beta N^{ih}(\omega))} \qquad (16)$$

The right hand side is also called the *energy mean.* Since $\ln(Z(\beta))$ is *convex* in Θ [6, 4], the gradient can be approximated by stochastic relaxation [4]:

Algorithm 1 (Hyperparameter Estimation)

① *Set $k = 0$, initialize $\widehat{\beta}^0$ and let $N^{ih}(\widehat{\omega})$ denote the number of inhomogeneous cliques in the estimate of the labeling.*

② *Using SA at a fixed temperature T, generate a new labeling η, sampling from*

$$P(\mathcal{X}=\omega) = \frac{\exp\left(-\frac{\widehat{\beta}^k}{T}\sum_{\{s,r\}\in\mathcal{S}}\delta(\omega_s,\omega_r)\right)}{Z(\widehat{\beta}^k)}. \qquad (17)$$

Compute the number of inhomogeneous cliques $N^{ih}(\eta)$ in η.

③ *If $N^{ih}(\eta) \approx N^{ih}(\widehat{\omega})$ then stop, else $k = k+1$. If $N^{ih}(\eta) < N^{ih}(\widehat{\omega})$ then decrease $\widehat{\beta}^k$, if $N^{ih}(\eta) > N^{ih}(\widehat{\omega})$ then increase $\widehat{\beta}^k$, and goto Step ②.*

Using Equation (14) and Algorithm 1, we can estimate the model parameters, *if* we have a realization of the label field. Unfortunately such a *labeled sample* is unknown in real-life applications, we have to use an estimation method which works with *unlabeled samples.* We have chosen a recent method, called Iterated Conditional Estimation (ICE) [2] to solve this problem.

Let us consider an estimator $\mathcal{E}_\Theta(\mathcal{F},\omega)$ of Θ (ML, for instance). Since realizations of the label field are unknown, the direct use of $\mathcal{E}_\Theta(\mathcal{F},\omega)$ is impossible, we have to approximate it. The best approximation, in the mean-square sense, is the *conditional expectation.* Since $E\{\mathcal{E}_\Theta \mid \mathcal{F},\omega\}$ depends on the parameters Θ, we need a parameter $\widehat{\Theta}^k$ previously defined by some way. This defines an iterative procedure [2]:

Algorithm 2 (ICE)

① *Set $k = 0$ and initialize $\widehat{\Theta}^0$.*

② *Generate n realizations (n is fixed a priori) $\widehat{\omega}^i (1 \le i \le n)$ of the label field based on $\widehat{\Theta}^k$.*

③ *Based on the sample $\widehat{\omega}^i (1 \le i \le n)$, $\widehat{\Theta}^{k+1}$ is obtained as the conditional expectation*

$$\widehat{\Theta}^{k+1} = E\{\mathcal{E}_\Theta \mid \mathcal{X}=\omega\} \approx \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_\Theta(\mathcal{F},\widehat{\omega}^i). \qquad (18)$$

④ *Goto Step ② until $\widehat{\Theta}$ stabilizes.*

There is still one problem to solve, namely the initialisation of the parameters in Step ①. We have used a simple algorithm to initialize the Gaussian parameters. The method, proposed by Postaire and Vasseur [7], consists of the geometrical analysis of the hystogram, regarded as a Gaussian mixture, in order to determine its modes. For the hyperparameter β, the initial value is not crucial, pratically any value between 0.5 and 1 is good.

The complete segmentation process is the following: Given an image $\mathcal{F}$, compute the histogram and initialize the mean and the deviation of the classes, using the geometrical analysis of the histogram. Then, using the ICE algorithm, estimate Θ. Once the final estimate $\widehat{\Theta}$ of the parameters is obtained, one proceeds to the ordinary segmentation with known parameters. The formulation of the unsupervised segmentation algorithm used for the simulations is the following:

Algorithm 3 (Unsupervised Segmentation)

① *Given an image $\mathcal{F}$, compute its histogram and for each $\lambda \in \Lambda$, initialize μ_λ and σ_λ using the histogram. β is initialized in an ad-hoc way.*

② **(Estimation)** *Using the ICE algorithm, get an estimate $\widehat{\Theta}$ of the parameters.*

③ **(Segmentation)** *Given the parameters $\widehat{\Theta}$, do an ordinary supervised segmentation to get the MAP estimate of the label field given $\mathcal{F}$ and $\widehat{\Theta}$.*

Parameter	Unsupervised		Supervised
	Initial	Final	
μ_0	83.5	84.3	85.48
σ_0^2	256.0	480.5	446.60
μ_1	100.0	117.3	115.60
σ_1^2	169.0	416.3	533.97
μ_2	152.5	148.1	146.11
σ_2^2	676.0	457.8	540.32
μ_3	181.5	178.5	178.01
σ_3^2	100.0	490.9	504.34
β	0.7	1.0	1.0

Table 1: Parameters of the synthetic image.

3. EXPERIMENTAL RESULTS

We have tested the proposed monogrid unsupervised algorithm on noisy synthetic and real images. The algorithm was implemented on a Connection Machine CM200. We have compared the obtained parameters and segmentation results to the supervised results. In general, the quality of unsupervised results are as good, or sometimes slightly better, than the results of supervised segmentation. For example, the number of misclassified pixels on the synthetic image was 112 (0.68%) with the supervised algorithm and 103 (0.63%) with the unsupervised method. We observed, however, that the unsupervised algorithm is more sensitive to noise than the supervised one. This is due to the initial conditions (in particular the initialization of the mean and the variance of the classes). For example, in the case of the synthetic image with SNR= $3dB$ one class has been lost. But with SNR= $5dB$, the result is as good as for the supervised algorithm.

Before evaluating the results, let us explain some important points of the implementation. The only parameter which has to be defined by the user is the number of classes (or regions). All the other parameters are estimated automatically from the data. For the hyperparameter, we have chosen $\beta = 0.7$ as initial value. Experiments show that the initial value is not vital, practically any value between 0.5 and 1 is good. In Step ② of Algorithm 3), we use the ICE algorithm (see Algorithm 2) to iteratively reestimate the parameters. We have chosen ICM [1] to generate labelings because of its rapidity: Given the parameters $\widehat{\Theta}^n$, the ICM is used to maximize the a posteriori probability of the label field ω. Suppose, that ICM converges in N iterations (N is typically less than 10) given N realizations of ω. Using these labelings, we have to compute N ML estimates of Θ. Once the sequence $\widehat{\Theta}^n$ becomes steady, the estimation step is finished and one proceeds to the segmentation (with known parameters) using the Gibbs sampler.

The algorithms were tested on a synthetic (Figure 1) and on a satellite image (Figure 2). We also give the histograms, since the initial estimates are based on them. In Table 1, we compare the parameters obtained by the unsupervised algorithm to the ones used for the supervised segmentation.

In Table 2, we give the computer time of the estimation and segmentation (VPR means the number of virtual processors per physical processor). As we can see, the estimation requires much more time than the segmentation. The hyperparameter estimation requires the largest part of the computer time since it consists of generating new labelings by Simulated Annealing in Step ② of Algorithm 1.

4. CONCLUSION

In summary, the presented unsupervised algorithms provide results comparable to those obtained by supervised segmentations, but they require much more computing time and they are slightly more sensitive to noise. The main advantage is, of course, that unsupervised methods are completely data-driven. The only input parameter is the number of regions. We believe that, for unsupervised methods, the main problem is still the initialization of the Gaussian parameters. Hence, a natural extension of this work would be to look for more efficient initialization techniques.

5. REFERENCES

[1] J. Besag. On the statistical analysis of dirty pictures. *Jl. Roy. Statis. Soc. B.*, 1986.

[2] W. Pieczynski. Statistical image segmentation. In *Machine Graphics and Vision, GKPO'92*, pages 261–268, Naleczow, Poland, May 1992.

[3] Z. Kato, J. Zerubia, and M. Berthod. Bayesian image classification using markov random fields. In A. Mohammad-Djafari G. Demoments, editor, *Maximum Entropy and Bayesian Methods*, pages 375–382. Kluwer Academic Publisher, 1993.

[4] D. Geman. Bayesian image analysis by adaptive annealing. In *Proc. IGARSS'85*, pages 269–277, Amherst, USA, Oct. 1985.

[5] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):799–813, Aug. 1989.

[6] R. J. Baxter. *Exactly Solved Models in Statistical Mechanics.* Academic Press, 1990.

[7] J. G. Postaire and C. P. A. Vasseur. An approximate solution to normal mixture identification with application to unsupervised pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(2):163–179, March 1981.

Image	VPR	Total CPU time	Estimation	Segmentation
Synthetic	2	249.75 sec.	237.00 sec.	12.75 sec.
Satellite	32	3576.58 sec.	3270.78 sec.	305.81 sec.

Table 2: Computer times.

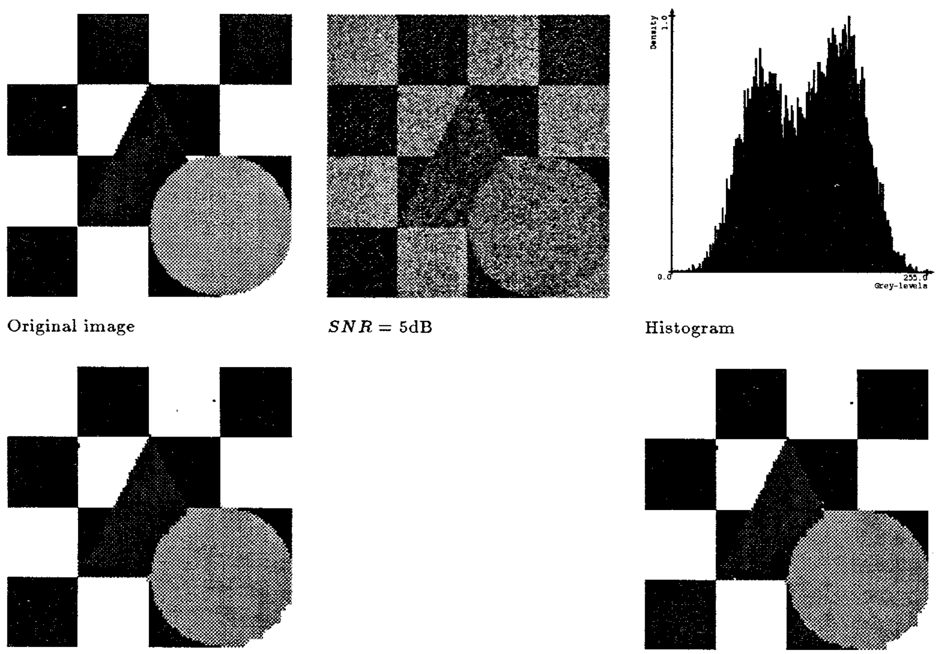

Figure 1: Supervised and unsupervised segmentation results with the Gibbs Sampler.

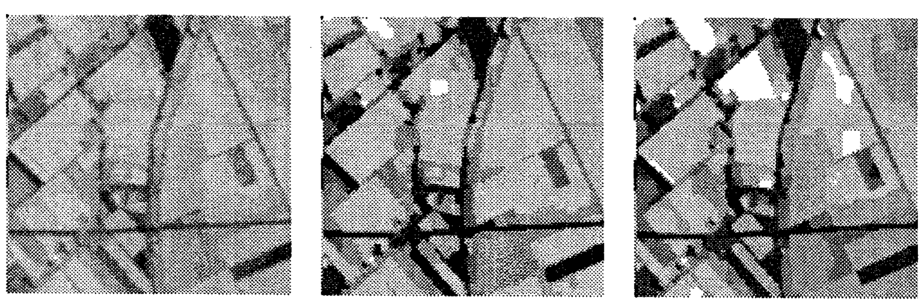

Figure 2: Supervised and unsupervised segmentation results with the Gibbs Sampler.