SATELLITE IMAGE CLASSIFICATION USING A MODIFIED METROPOLIS DYNAMICS

Z. Kato and **J. Zerubia** and **M. Berthod**
INRIA, Sophia-Antipolis, 2004 Route des Lucioles, 06565 Valbonne France

ABSTRACT

In this paper, we present a pseudo-stochastic variation of the Metropolis dynamics for combinatorial optimization in image classification using Markov Random Fields. At high temperature, the behavior of our algorithm is similar to the stochastic ones. However, if the temperature is less than a certain threshold, it becomes deterministic. The "length" of the "pseudo-stochastic" phase is controlled by a constant threshold used in the modified dynamics. The algorithm yields an approximate but usually good solution to the optimization problem. The algorithm runs on a Connection Machine. It is applied to the standard pixel classification problem; objective and subjective comparisons with other algorithms have been made.

INTRODUCTION

Markov Random Fields (MRF) have become more and more popular during the last few years in image processing. A good reason for that is that such a modelization is the one which requires the less a priori information on the world model. Many standard computer vision problems, specifically early vision ones such as image classification [1], can thus be expressed quite naturally as combinatorial optimization problems. Many heuristics have been proposed to solve them : Iterated Conditional Modes (ICM) [2, 3], Graduated Non-Convexity [4, 5], Mean Field Annealing [6, 7], Simulated Annealing [8, 9, 10] ...

We propose here a modified version of the Metropolis algorithm [10] called Modified Metropolis Dynamics (MMD). This methode have been implemented on a Connection Machine CM2 and compared to ICM, Metropolis algorithm and Gibbs sampler on noisy synthetic data as well as SPOT images.

MODIFIED METROPOLIS DYNAMICS

Herein, we are interested in the following general problem: we are given a set of sites $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ each of which may belong to any one of L classes (or equivalently take any label from $\Lambda = \{1, 2, \ldots, L\}$). We are also given a MRF on these sites, defined by the so-called clique potentials and by a neighborhood-system $\mathcal{G}$. Let C denote a clique of $\mathcal{G}$, and $\mathcal{C}$ the set of all cliques. Also $C_s = \{C : s \in C\}$.

A global discrete labeling ω assigns one label ω_s ($\omega_s \in \Lambda$) to each site s in $\mathcal{S}$. The restriction of ω to the sites of a given clique C is denoted by ω_C. The definition of the MRF is completed by the knowledge of the clique potentials $E_C(\omega_C)$ for every C in $\mathcal{C}$ and every ω in Ω, where Ω is the set of all possible L^N discrete labelings.

The Model

Given $\mathcal{F} = \{f_s\}_{s \in \mathcal{S}}$ a set of image data where f_s stands for the greylevel at pixel s. A very general problem is to find the labeling $\hat{\omega}$ which maximizes $P(\omega \mid \mathcal{F})$. Bayes theorem tells us that $P(\omega \mid \mathcal{F}) = \frac{1}{P(\mathcal{F})} P(\mathcal{F} \mid \omega) P(\omega)$. Actually $P(\mathcal{F})$ does not depend on the labeling ω and we have the assumption that $P(\mathcal{F} \mid \omega) = \prod_{s \in \mathcal{S}} P(f_s \mid \omega_s)$. It is then easy to see that the global labeling which we are trying to find is given by:

$$\hat{\omega} = \max_{\omega \in \Omega} \prod_{s \in \mathcal{S}} P(f_s \mid \omega_s) \prod_{C \in \mathcal{C}} \exp(-E_C(\omega_C)) . \qquad (1)$$

It is obvious from this expression that the a posteriori probability also derives from a MRF. The energies of cliques of order 1 directly reflect the probabilistic modeling of labels without context, which would be used for labeling the pixels independently. Let us assume that $P(f_s \mid \omega_s)$ is Gaussian, the class $\lambda \in \Lambda$ is represented by its mean value μ_λ and its deviation σ_λ. We get:

$$\begin{aligned} \hat{\omega} \;\; = \;\; & \min_{\omega \in \Omega} \left(\sum_{s \in \mathcal{S}} \left(\log(\sqrt{2\pi}\sigma_{\omega_s}) + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma^2_{\omega_s}} \right) \right. \\ & \left. + \sum_{C \in \mathcal{C}} E_C(\omega_C) \right) \qquad (2) \end{aligned}$$

Using the above equation, it is easy to define the local energy of any labeling ω at site s:

$$\mathcal{E}_s(\omega) = \log(\sqrt{2\pi}\sigma_{\omega_s}) + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma^2_{\omega_s}} + \sum_{C_s} E_{C_s}(\omega_{C_s}) \qquad (3)$$

The estimation of $\hat{\omega}$ is done through the energy minimization using a relaxation method with a Modified Metropolis Dynamics (MMD)[11].

The Algorithm

Herein we present a modified version of the Metropolis dynamics [10]. The difference between the original method and our approach is the choice of the threshold ξ used in the dynamics. For the original method, ξ is chosen randomly at each iteration, however for our algorithm, ξ is a constant threshold, say α, chosen at the beginning of the algorithm. This simply means that the jump to a new labeling state η is allowed if this does not increase "excessively" the energy. The threshold α controls this increase of energy. The algorithm is highly parallel and can be described as follows:

1. In order to get the convergence of the algorithm, partition the entire image into disjoint regions $\mathcal{R}_n$ $(1 \leq n \leq M)$ such that pixels which belong to the same region are conditionally independent given the data of all the other regions. Pick up randomly an initial configuration ω^0, with $k = 0$ and $T = T_0$.
2. Using a uniform distribution, pick up a global state $\eta \in \Omega \setminus \{\omega^k\}$. For each site $s \in \mathcal{R}_n$ $(1 \leq n \leq M)$, the local energy $\mathcal{E}_s(\eta')$ is computed using Eq. (3) with $\eta' = [\omega^k_{s_1}, \omega^k_{s_2}, \ldots, \eta_s, \ldots, \omega^k_{s_N}]$. Calling $\Delta\mathcal{E}_s = \mathcal{E}_s(\eta') - \mathcal{E}_s(\omega^k)$, the new label η_s at site s is accepted if $\alpha \leq \exp\left(-\frac{\Delta\mathcal{E}_s}{T}\right)$, where $\alpha \in [0,1]$ is the constant threshold. Otherwise $\omega^{k+1}_s = \omega^k_s$.
3. Decrease of the temperature $T = T_{k+1}$ (k = number of iterations) and goto 2 if $\Delta\mathcal{E}_{glob} > threshold$.

If the temperature is less than a certain threshold $(= \frac{\Delta\mathcal{E}_{min}}{-\log\alpha}$, see [11]) then only the jumps to states of lower energy are allowed. The algorithm converges to a local minimum (see [11] for a complete proof).

There is no explicit formula to get the treshold α. In practice, α is determined empirically according to the following rule: in the case of a noisy image, α is chosen nearly equal to zero; otherwise α is chosen nearly equal to one.

EXPERIMENTAL RESULTS

All the algorithms used for the simulation have been implemented on a Connection Machine CM2 [12] with 8K processors.

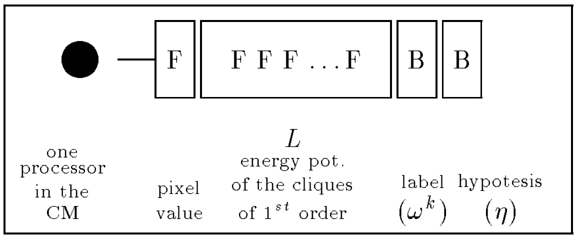

Figure 1: Memory complexity on the CM (Field sizes: B=8 bits and F=32 bits)

Complexity

In the general case of relaxation with L classes, the algorithm will associate L floating numbers $E_{\{s\}}(1), \ldots, E_{\{s\}}(L)$ to pixel s (recall that $E_{\{s\}}(\lambda)$ denotes the first order clique-potential). These data with all other data associated with pixel s (the greylevel, the label (ω^k_s) and the hypotesis (η_s)) are gathered in one (virtual) processor in the CM as shown in Fig. 1. Thus the memory complexity is $LF + F + 2B$.

At each iteration, random numbers are generated for the setup of the hypotesis η and the energy change $\Delta\mathcal{E}_s$ is computed. It is then easy to see that, in the parallel case, the computer-time complexity of a complete sweep is $VPR \cdot M(\epsilon + \tau)$; calling VPR the Virtual Processor Ratio of the given machine, ϵ the computer-time needed for the calculation of $\Delta\mathcal{E}_s$ and τ the one needed to generate a random number (recall that M is the number of regions). In the sequential case, it equals to $N(\epsilon + \tau)$. One should mentioned that, in all cases, the computer-time is independent on the number of classes.

The total speed-up depends on the size of the neighborhood of the MRF and the virtual processor ratio of the given machine. In the theory, we can have N/M speed-up if we have enough processors.

Comparison with Other Methods

Now, we present the results obtained on a noisy checkerboard image (128×128) with a SNR= -5dB with 2 classes (see Fig. 2, Tables 1, 2) and on a SPOT image (256×256) with 4 classes (see Fig. 3, Table 1). For the simulations, the following parameters have been used:

- *Temperature*: For ICM the temperature is constant. For the other algorithms, we have used an exponential schedule: $T_{k+1} = 0.95T_k$. The initial temperature is $T_0 = 10$ for the synthetic image and $T_0 = 4$ for the SPOT image.
- *Mean and deviation of each class*: They are computed using a supervised learning method.

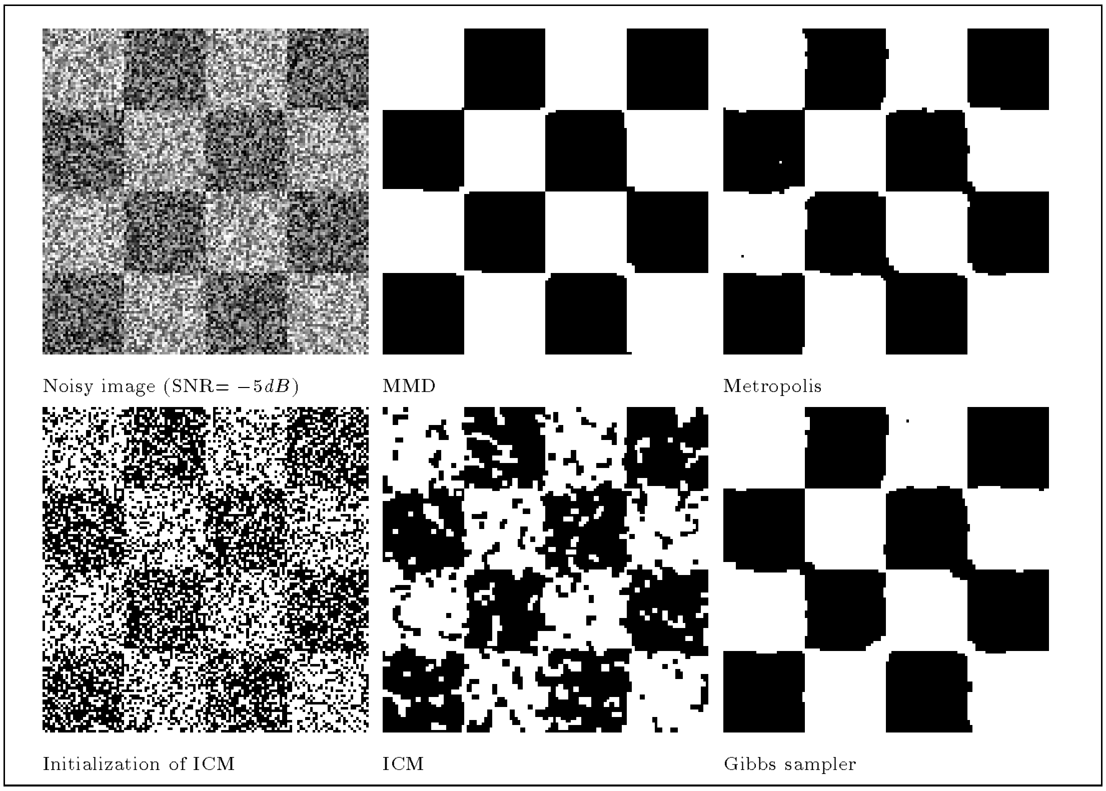

Figure 2: Results with the checkerboard image with 2 classes

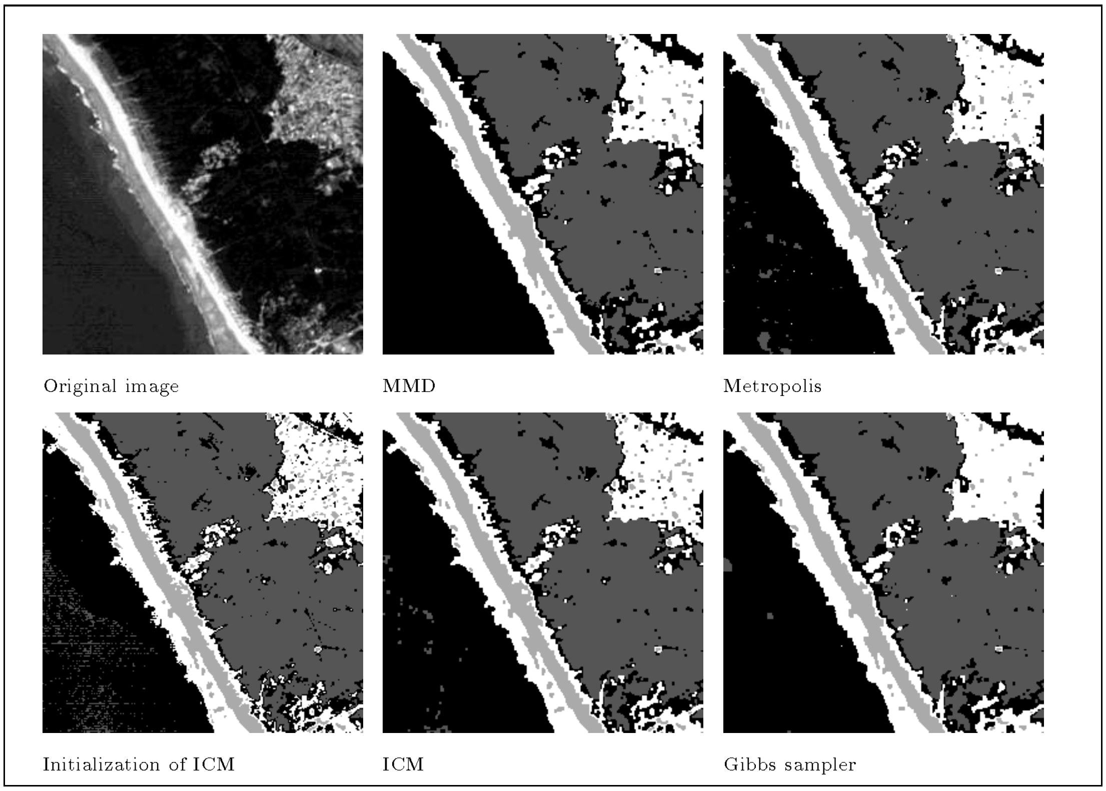

Figure 3: Results with the SPOT image with 4 classes

A	VPR	Nb. of Iter	Total time	Time per It.
ICM	2	9	0.249 sec.	0.027 sec.
Metrop.	2	90	26.32 sec.	0.293 sec.
Gibbs	2	86	48.51 sec.	0.564 sec.
MMD	2	79	2.37 sec.	0.030 sec.

B	VPR	Nb. of Iter	Total time	Time per It.
ICM	8	6	0.875 sec.	0.146 sec.
Metrop.	8	52	61.39 sec.	1.181 sec.
Gibbs	8	57	235.68 sec.	4.135 sec.
MMD	8	35	6.28 sec.	0.179 sec.

Table 1: Computer times. A: on the checkerboard image with 2 classes, B: on the SPOT image with 4 classes

- *Choice of the clique-potentials*: We have used second order clique-potentials. If two neighboring sites have different labels, the energy is 2.0 for the synthetic image and 0.5 for the other one. If they have the same label, the energy is 0.2 (resp. -0.5).
- *Initialization of the labels*: Random values are assigned to the labels for the initialization of Gibbs, Metropolis and MMD. As for ICM, the initial labels are obtained using only the Gaussian term in Eq. (2).
- *Initialization of α for MMD*: $\alpha = 0.1$ for the checkerboard and $\alpha = 0.8$ for the SPOT image.

	win	alert	error
ICM	0.8439	0.1561	0.1639
Metropolis	0.9902	0.0098	0.0156
Gibbs	0.9866	0.0134	0.0194
MMD	0.9989	0.0011	0.0083

$$\begin{aligned} win &= \frac{Number\, of\, white\, pixels\, classed\, white}{Number\, of\, input\, white\, pixels} \\ alert &= \frac{Number\, of\, white\, pixels\, classed\, black}{Number\, of\, input\, white\, pixels} \\ error &= \frac{Number\, of\, black\, pixels\, classed\, white}{Number\, of\, input\, black\, pixels} \end{aligned}$$

Table 2: Objective performance comparison on the checkerboard image

CONCLUSION

In this paper, we have proposed a suboptimal algorithm applied to image classification. This technique compares favorably with the classical relaxation algorithms (ICM, Metropolis, Gibbs sampler) giving good results and being quite fast on a parallel machine.

References

[1] P. Chou and C. Brown. The theory and practice of bayesian image labeling. *International Journal of Computer Vision*, 4:185–210, 1990.

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

[3] F. C. Jeng and J. M. Woods. Compound Gauss - Markov Random Fields for Image Estimation. *IEEE Trans. Acoust., Speach and Signal Proc.*, ASSP-39:638–697, 1991.

[4] A. Blake and A. Zisserman. Visual reconstruction. *MIT Press, Cambridge - MA*, 1987.

[5] A. Rangaraja and R. Chellappa. Generalised Graduated Non-Convexity algorithm for Maximum A Posteriori image estimation. *Proc. ICPR*, pages 127–133, Jun. 1990.

[6] D. Geiger and F. Girosi. Parallel and deterministic algorithms for MRFs : surface reconstruction and integration. In *Proc. ECCV90*, Antibes, France, 1990.

[7] J. Zerubia and R. Chellappa. Mean field approximation using compound Gauss-Markov Random field for edge detection and image restoration. *Proc. ICASSP, Albuquerque, USA*, 1990.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[9] P.J. Van Laarhoven and E.H. Aarts. Simulated annealing : Theory and applications. *Reidel Pub., Dordrecht, Holland*, 1987.

[10] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. of Chem. Physics, Vol. 21, pp 1087-1092*, 1953.

[11] Z. Kato. Modélisation Markovienne en classification d'image. mise en oeuvre d'algorithmes de relaxation. *Rapport de stage de DEA, Université de Nice*, 1991.

[12] W. D. Hillis. *The Connection Machine.* Cambridge, MA, 1985.