Multi-Temperature Annealing: a New Approach for the Energy-Minimization of Hierarchical Markov Random Field Models

Josiane ZERUBIA, Zoltan KATO and 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 - email: <*name*>`@sophia.inria.fr`

Abstract

As it is well known, optimization of the energy function of Markov Random Fields is very expensive. Hierarchical models have usually much more communication per pixel than monogrid ones. This is why classical annealing schemes are too slow, even on a parallel machine, to minimize the energy associated with such a model. However, taking benefit of the pyramidal structure of the model, we can define a new annealing scheme: the Multi-Temperature Annealing (MTA), which consists of associating higher temperatures to coarser levels, in order to be less sensitive to local minima at coarser grids. The convergence to the global optimum is proved by a generalisation of the annealing theorem of Geman and Geman. We have applied the algorithm to image classification and tested it on synthetic and real images.

1 Introduction

There are many approaches in multigrid image models. Bouman [2] proposes a multiscale MRF model, where each scale is causally dependent on the coarser grid field above it. Heitz et al. in [7] propose a consistent multiscale MRF model for motion analysis. We have applied this model to image classification and proposed a new hierarchical Markov Random Field (MRF) model in [5, 4]. Herein, we are interested in efficient energy-minimization methods for hierarchical MRF models.

In the following (for more details on the model, see [5, 4]), we suppose that an image is a $W \times H$ lattice, so that $\mathcal{S} = \{(i,j) : 1 \leq i \leq W \text{ and } 1 \leq j \leq H\}$, and $W = w^n$, $H = h^m$. First, we generate a label pyramid but we keep the whole observation field: For all $1 \leq i \leq M$ ($M = \inf(n,m)$), $\mathcal{S}$ is divided into blocks of size $w^i \times h^i$. These blocks will form a coarser scale.

*This work has been partially funded by CNES and DRED/GdR-TdSI

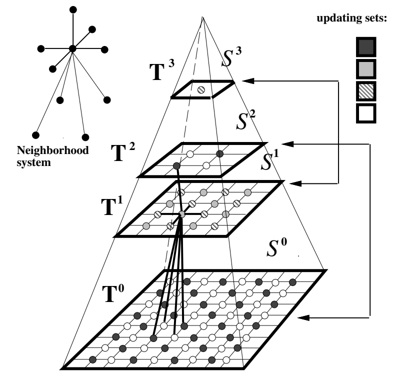


Figure 1: Neighborhood system and relaxation scheme on the pyramid. The levels connected by pointers are updated at the same time.

The labels assigned to the sites of a block are supposed to be the same over the whole block. Then, a block is reduced to a unique site at the corresponding level in the pyramid. We have the same neighborhood structure at coarser grids in the pyramid as on the the finest (initial) grid (we note the cliques at level i by C^i). We introduce new interactions between two neighbor grids in the pyramid (see Figure 1). The energy function is of the following form:

$$U(\omega) = \sum_{i=0}^{M} \sum_{C^i \in \mathcal{C}^i} V^i_{C^i}(\omega) + \sum_{C^* \in \mathcal{C}^*} V_{C^*}(\omega)$$

where C^* denotes the new cliques siting astride two neighbor grids and $V^i_{C^i}(\omega)$ denotes the potentials at level i, which are derived by simple computation from the potentials of the finest grid (for more details see [4]).

2 Multi-Temperature Annealing

Now, let us see how the energy $U(\omega)$ is minimized. If we use a deterministic relaxation method where the temperature parameter is kept constant during the iterations (for example ICM [1]), then the original formulation of the algorithm does not change. The only difference is that we work on a pyramid and not on a rectangular shape as in the monogrid case. We can easily parallelize this algorithm using the coding technique described by Besag in [1]: we partition the pyramid into disjoint updating sets so that pixels which belong to the same set are conditionally independent, given the data of all the other sets. This enables us to update different levels at the same time (see Figure 1). Let us consider in the following a relaxation algorithm where the temperature change during the iterations (Simulated Annealing [3], for example). The temperature-change is controlled by the so-called *annealing schedule*. We have three ways of annealing. The first two are well known [6], they require no modification of the original algorithm, except that we work on a pyramid instead of a rectangular shape. The third one is a new annealing schedule: the Multi-Temperature Annealing which is the most efficient with the hierarchical model:

1. *Homogeneous annealing:* We assign to each level of the pyramid the same, initially high, temperature. The relaxation is performed with this fixed temperature until an equilibrium is reached The temperature is then lowered. The algorithm is stopped at a temperature closed to 0.

2. *Inhomogeneous annealing:* The same initially high temperature is assigned to each level, however the temperature is now lowered after each transition.

3. *Multi-Temperature Annealing (MTA):* The basic idea here is to associate higher temperatures to coarser levels in the pyramid which enables to be less sensitive to local minima. However at a finer resolution, the relaxation is performed at a lower temperature (at the bottom level, it is closed to 0). For the cliques siting between two levels, we use either the temperature of the finer level or the one of the coarser level (but once chosen, we always keep the same choice throughout the algorithm).

For the parallelization, we use the same coding technique as in the deterministic case. In all cases, the final configuration of the finest level is taken as the solution of the problem. The following theorem assures the convergence to a global optimum using the MTA schedule: Let π_0 be the uniform distribution on the set of globally optimal configurations, and define $U^{sup} = \max_\omega U(\omega)$, $U^{inf} = \min_\omega U(\omega)$ and $\Delta = U^{sup} - U^{inf}$. Furthermore, let us suppose that the sites are visited for updating in the order $\{n_1, n_2, \ldots\} \subset \mathcal{S}$. The following theorem gives an annealing schedule, basically the same as in [3]. *However, the temperature here is a function of the iteration k and $C \in \mathcal{C}$.*

Theorem 1 (Multi-Temperature Annealing) *Assume that there exists an integer $\kappa \geq N$ such that for every $k = 0, 1, 2, \ldots$, $\mathcal{S} \subseteq \{n_{k+1}, n_{k+2}, \ldots, n_{k+\kappa}\}$. Let $T(k, C)$ be any decreasing sequence of temperatures for which*

1. *There exists T_k^{inf} such that for all $C \in \mathcal{C}$: $T_k^{inf} \leq T(k, C)$.*

2. *$\forall C:\ \lim_{k\to\infty} T(k, C) = 0$*

3. *For all $k \geq k_0$, for some integer $k_0 \geq 2$: $T_k^{inf} \geq N\Delta/\log k$.*

Then for any starting configuration $\eta \in \Omega$ and for every $\omega \in \Omega$:

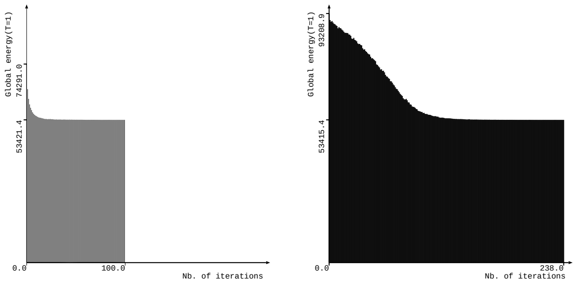
$$\lim_{k\to\infty} P(X(k) = \omega \mid X(0) = \eta) = \pi_0(\omega).$$

The first condition states the existence of a lower bound for the temperature function. The second condition is that the temperature decreases as the system evolves. The third condition assures that each "local" temperature decreases in an appropriate way, i.e. not too quickly. The complete proof of this theorem can be found in [4].

Now, we study the complexity of the optimization of the hierarchical model in terms of the required memory (or number of processors in the parallel implementation) and the required communication number compared to the monogrid model. Let us suppose that our image is of size $W \times H$ and our pyramid contains $M + 1$ levels. Without loss of generality, we can assume that $W/w \leq H/h$, where $w \times h$ is the block size, with both $w \geq 2$ and $h \geq 2$. The hierarchical model requires a maximum of $(1+1/w)WH$ processors, since all levels must be stored at the same time. The memory (or number of processors) required for the storage of these levels considering a rectangular shape, is given by:

$$WH \sum_{i=0}^{M} \frac{1}{(wh)^i} < \left(1 + \frac{1}{w}\right) WH$$

Considering only the first and second order cliques (mostly used in practice), it is clear that we need



MTA Inhomogeneous

Figure 2: Energy decrease with the MTA and inhomogeneous annealing schedule

$(wh+1)$ more communications per processor. Each site interacts with its ancestor (there is one) and its descendants (there are wh).

It turns out that the hierarchical model requires more processors and more computing time than a monogrid model. However, as we have shown in [4], experiments show that this model yields faster convergence (with respect to the number of iterations) for the stochastic relaxation algorithms and gives estimates which are closer to the global optimum.

3 Experimental Results

In Figure 3, we compare the inhomogeneous and MTA schedules on a noisy synthetic image using the Gibbs sampler. In both cases, the parameters were strictly the same, the only difference is the applied schedule. In Figure 2, we show the global energy (computed at a fixed temperature on the finest level) versus the number of iterations of the inhomogeneous and MTA schedules. Both reach pratically the same minimum (53415.4 for the inhomogeneous and 53421.4 for the MTA), however the inhomogeneous schedule requires 238 iterations (796.8 sec. CPU time) but the MTA schedule requires only 100 iterations (340.6 sec. CPU time) for the convergence. Other tests can be found in [4].

4 Conclusion

We have proposed a new annealing scheme, the Multi-Temperature Annealing, which is especially well adapted to the minimization of the energy function of

SNR= $3dB$ Inhomogeneous MTA

Figure 3: Results of the Gibbs sampler on a synthetic image with inhomogeneous and MTA schedules

hierarchical MRF models. This algorithm can be run in parallel on the entire pyramid and usually decreases the computing time compared to the classical schemes. A generalization of the annealing theorem of Geman and Geman [3] has been proposed, which gives a theoretical background for the convergence of this method towards a global optimum. We remark that MTA is a quite general annealing scheme. The theoretical study given in [4] is not limited to a pyramidal structure nor to a hierarchical MRF model.

References

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

[2] C. Bouman. A Multiscale Image Model for Bayesian Image Segmentation. Technical Report TR-EE 91-53, Purdue University, 1991.

[3] 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.

[4] Z. Kato, M. Berthod, and J. Zerubia. A hierarchical markov random field model and multitemperature annealing for parallel image classification. Research Report 1938, INRIA, Aug. 1993.

[5] Z. Kato, M. Berthod, and J. Zerubia. Multiscale markov random field models for parallel image classification. In *Proc. ICCV*, Berlin, May 1993.

[6] P. V. Laarhoven and E. Aarts. Simulated annealing : Theory and applications. *Reidel Pub., Dordrecht, Holland*, 1987.

[7] F. Heitz, P. Perez and P. Bouthemy. Multiscale Minimization of Global Energy Functions in Some Visual Recovery Problems. In *CVGIP: Image Understanding*, Vol. 59, No. 1. pp. 125–134, 1994.