

Markov random fields on a SIMD machine for global region labelling

Greg Budzban

Martin Marietta Electronic Systems, MP 1304, Orlando, FL 32862-8007

John DeCatrel

University of Central Florida, Department of Computer
Engineering, Orlando, FL 32816-0450

ABSTRACT

The Markov Random Field (MRF) formulation allows independence over small pixel neighborhoods suitable for SIMD implementation. The equivalence between the Gibbs distribution over global configurations and MRF allows describing the problem as maximizing a probability, or equivalently, minimizing an energy function (EF).

The EF is a convenient device for *integrating* "votes" from disparate, pre-processed features—mean intensity, variance, moments, etc. Contributions from each feature are simply weighted and summed. The EF is flexible and can be easily modified to capture *a priori* beliefs about the distribution of the configuration space, and still remain theoretically sound. A unique formulation of the EF is given. Notably, a deterministic edge finder contributes to the EF. Weights are independently assigned to each feature's report (indicators).

Simulated annealing is the theoretical mechanism which guarantees convergence in distribution to a global minimum. Because the number of iterations is an exponential function of time, we depart from theory and implement a fast, heuristic "cooling" schedule. A videotape of results on simulated FLIR imagery demonstrates real-time update over the entire image. Actual convergence is still too slow for real-time use ($O(1 \text{ min.})$), but the quality of results compares favorably with other region labelling schemes.

INTRODUCTION

For some tasks such as automatic terrain following or obstacle avoidance, central to the design of associated automatic vision systems are techniques that attempt to accurately label regions. Unfortunately, classical segmentation techniques have proven to lack stability and robustness. Sometimes the smallest variation in environmental conditions and the resulting change in image intensity will cause drastic differences in the output of classical segmenters and region labellers. Many of these segmentation algorithms are *ad hoc* with little or no mathematical justification. Markov random

copyright ©1991 Martin Marietta Corp. All rights reserved.

field theory applied to the segmentation problem provides a possible alternative to these classical approaches.

The application of Markov random field (MRF) techniques in image processing applications has become fairly widespread.³ Most of the work, however, remains in the academic arena on binary images. The work done on images with, for example, 256 grey levels requires such extensive computation time for convergence to the optimal solution that it renders MRFs useless for real world problems. This paper will discuss attempts to use MRFs to solve general region labeling problems in a real time setting. These attempts made use of Martin Marietta's Geometric Arithmetic Parallel Processor (GAPP)TM, a SIMD array with over 40,000 processing units. Some of the algorithmic techniques to be discussed reflect the requirement of keeping all the data in local RAM (128 bits per processor) to avoid spending time moving images in and out of the SIMD array.

The outline of this paper will be as follows. The first section provides an overview of the history and mathematical background of MRFs including the role of simulated annealing in the solution of the problem. Next, a discussion of the constraints that real time requirements impose on the system will be followed by an exposition of some of the techniques used to overcome these constraints. Then, the particular algorithm used to program the GAPP array is presented. Finally, we discuss some results and conclude by pointing out some areas requiring further investigation.

2. MATHEMATICAL BACKGROUND OF MRFs

Random field techniques have been studied in various forms since Gibbs and others began investigating statistical mechanics late in the last century. Gibbs random fields, of which the Ising model is perhaps the simplest, were studied for a variety of reasons. Ising and others sought to study the concept of phase transitions in ferromagnetism and other complex systems. They utilized a model of the system which computed the probabilities of various configurations on the basis of a local potential function. In other words, the probability of a particular configuration of charged particles in a lattice, w , could be expressed in the following form. If D represented the possible lattice sites, and x, y were elements of D then,

$$P(w) = \frac{1}{z} \exp \left[\sum_{x,y \in D} w(x)w(y)U(x,y) \right]$$

where $U(x, y)$ was the potential function, $w(x) = 1$ if site x was occupied (or positive) and 0 otherwise, and z was a normalizing constant used to ensure that P was a probability.

GAPPTM is a trademark of Martin Marietta Corp.

Another class of random fields was introduced by Dobrushin. His theory was an attempt to extend the notion of a one dimensional time indexed Markov process to one that was indexed spatially. Recall that a Markov process is governed by conditional probabilities that say: the probability that the process is in a particular state depends only on the state of the process in the previous time increment. That is at time k ,

$$P(X_k = x_k | X_j = x_j, j = 1, \dots, k-1) = P(X_k = x_k | X_{k-1} = x_{k-1}).$$

Dobrushin generalized this to a spatial process on a lattice of sites where the state of a given node of the lattice depended only on the states of the nearest spatial neighbors. In other words if s is the site in question, G_s is the set of the nearest spatial neighbors of s , then

$$P(X_s = x_s | X_r = x_r, r \neq s) = P(X_s = x_s | X_r = x_r, r \in G_s).$$

While this generalization was interesting, its usefulness was limited since the controlling conditional probabilities were virtually impossible to determine.

The key breakthrough occurred when in the late 1960's Clifford and Hammersley, Spitzer⁶, and others proved that these two seemingly distinct formulation of random fields were in fact equivalent. This theorem allowed the computation of the local conditional probabilities by utilizing the potential function that characterized the Gibbs distribution.

3. SIMULATED ANNEALING

Much of the current interest in MRF approaches stems from the 1984 article by Geman and Geman¹ which detailed techniques to generate the maximum *a posteriori* (MAP) estimate of the original image given a degraded image using MRF techniques. They mention in this paper the connection between restoring degraded images and the segmentation of imagery. Their approach utilized local (i.e. pixel) changes in the estimate to create a sequence of images that would generate the MAP solution. These results were intriguing both for the techniques themselves and for the level of mathematical rigor established in the paper.

Perhaps, Geman and Geman's main contribution (besides the mathematical framework) was stochastic relaxation. Stochastic relaxation was the term they used to describe the simulated annealing optimization technique that was central to determining the MAP solution. Their scientific contribution included theorems which proved that given any initial configuration of states, their algorithm would converge in distribution to the MAP estimate.

Simulated annealing as an optimization procedure introduces a parameter T , referred to as temperature, that controls the convergence of the system. The technique seeks to overcome the inadequacies of gradient descent type optimization procedures by allowing random uphill moves in the energy landscape to avoid falling into a local

minimum and converging to a non-optimal solution. Another contribution of the paper was a rigorous result concerning a sequence of temperatures $T_{(1)}, T_{(2)} \dots$, called an annealing schedule, that assured convergence. Their result was the following:

Let $T(t)$ be any decreasing sequence of temperatures for which

a) $T(t) \rightarrow 0$ as $t \rightarrow \infty$

b) $T(t) > \frac{N\Delta}{\ln(t)}$, where N is the number of pixels in the image, Δ a constant

then for any starting configuration, the sequence of images converges in distribution to the MAP estimate.

Unfortunately, there is no practical way this schedule can be kept (Notice , for example, that to reach a temperature of 1.0, the number of iterations must reach $\exp(N)$). Various "fast" annealing schedules have since been proposed, but even these require far too much computation time to be effective for real time processing. In the next section, we will discuss the simplifications that were required in the annealing schedule to approach processing times on the order of one minute per image.

Real time processing imposes other requirements on an automatic vision system developed for a non-laboratory environment. As indicated before, Martin Marietta's GAPP processor was designed with specific size constraints in mind. Because of the lack of local memory, approximations were made to stochastic relaxation to optimize processor speed. One of these approximations was the change from floating point calculations to integer arithmetic in the following way. A powerful feature of stochastic relaxation, in fact any simulated annealing based optimization routine, is its ability to accept changes which actually increase the energy of the system (or equivalently, decrease the probability of the current configuration). Thus to determine a configuration of states w^* such that

$$P(w^*) = \frac{1}{z} \exp(-U(w))$$

is at a maximum, the following procedure is implemented.

Step 0. Let w_0 be a random initial configuration of states and let $T(t)$, $t = 1, 2, \dots$ be an annealing schedule. (i.e. $T(1)$ is "large" and $T(t)$ approaches zero as $t \rightarrow \infty$).

Step 1. Choose a pixel, p_{ij} , and propose a state change from s_0 to s_1 .

Step 2. Compute a uniform random number $\beta(t)$, $0 < \beta(t) < 1$.

Step 3. Let $q = \frac{P(s_1)}{P(s_0)}$. Then

$$P_{ij} = \begin{cases} s_1 & \text{if } q > \beta(t) \\ s_0 & \text{otherwise} \end{cases}$$

Step 4. Let $t=t+1$ and go to step 1.

This procedure requires floating point arithmetic in a number of steps. To implement this procedure on the GAPP, the following approximation was used. Let

$$P(s_1) = e^{-\alpha_1/T(t)}$$

$$P(s_0) = e^{-\alpha_0/T(t)} \quad \text{for some } \alpha_1, \alpha_0$$

Then the algorithm says to change s_0 to s_1 if

$$\beta(t) < \frac{P(s_1)}{P(s_0)} = \frac{e^{-\alpha_1/T(t)}}{e^{-\alpha_0/T(t)}} = e^{(\alpha_0 - \alpha_1)/T(t)}$$

Thus the change occurs if

$$\ln \beta(t) < \frac{\alpha_0 - \alpha_1}{T(t)}$$

$$\text{or } T(t) \ln \beta(t) < \alpha_1 - \alpha_0 \quad \text{where } 0 < \beta(t) < 1$$

$$\Rightarrow -\ln \beta(t) T(t) > \alpha_1 - \alpha_0$$

$$\text{or } \ln \left(\frac{1}{\beta(t)} \right) T(t) > \alpha_1 - \alpha_0$$

Now $\beta(t)$ is uniform in $(0,1)$, thus

$$\text{prob} \left(\ln \left(\frac{1}{\beta(t)} \right) \in (0,5) \right) \approx 1.$$

Thus change from s_0 to s_1 if $\alpha_1 - \alpha_0 < T'(t)$ where $T'(t) = T(t) \ln \left(\frac{1}{\beta(t)} \right)$. This shows that one can work strictly with the potential functions and avoid taking exponentials.

4. ALGORITHM

A key idea in implementing our algorithm is: if a region label is the same as its neighbor change an "energy" value associated with that label site. Our goal is to lower the energy if both

- a label and its neighbor is the same
- the difference (disparity) between a site's feature and its neighbor is low

Conversely, raise the energy at sites where labels are the same but feature disparity is high. Therefore, over time, improved labelling is indicated by lower energies over the lattice.

Neighbors are defined over some regular, local neighborhood as further described in the detailed algorithm. Because of MRF locality many sites can be updated simultaneously, as long as a site and its immediate neighbors are not updated at the same time. By alternating sites considered for update and sites considered as neighbors

(no update), all site labels and energies are allowed to change. Over time this has the effect of a relaxation process. The influence of site energies is propagated via neighbors throughout the lattice.

Image metrics or features include a 5 x 5 intensity mean image, a standard deviation image, and an edge image derived from a simplified Canny edge detector. The incorporation of these features is highly modular, and other features can easily replace these in our algorithm. The detailed algorithm now follows. // indicate comments.

```
// This is a one-time set of initializing routines; initialization:
```

```
randomize label plane
extract features:
    5x5 mean filter image  $\mu = f_1$ 
    std deviation  $\sigma = f_2$ 
    Kirsch 12-directional edge image
obtain average disparities  $\overline{\Delta f_1}$  and  $\overline{\Delta f_2}$ 
initialize control parameter,  $T \in 20 \dots 30$ 
```

```
// The toggling of masks allows all sites that are not neighbors to be updated
concurrently. Sites take turns being neighbors. Because neighbors are arranged
in cross pattern five pixels away and one pixel away, two masks are used. Odd and
even indicate i+j matrix positions as usual.
```

```
iteration (T constant)
    generate proposed new label  $l^+$ 
    enable even 5 x 5 regions
    enable even pixels
    call kernel
    toggle mask to odd pixels
    call kernel
    toggle mask to odd 5 x 5 regions
    toggle mask to even pixels
    call kernel
display iteration results ( $l$ ,  $E(l)$ )
```

```
// This is the top-most control structure. The control parameter or temperature
is decremented down to zero. This is a fast cooling schedule since T reflects
the log of the number of prescribed iterations.
```

cooling

```
repeat until  $T = 0$ 
  repeat until (at least GAPP plane width)
    call iteration
   $T = T - 1$ 
```

// a clamped linear function is used to map disparities to energies. Desirable outputs of this function are that its range should have small integral values for simplicity but be broad enough to capture significant change in disparity. Clamping erases large perturbations that are not significant.

kernel

```
for all 8 neighbors do
  for  $f_1$  and  $f_2$  do
    calculate disparity  $\Delta f$  between pixel and neighbor:
       $\Delta f = |f_i(\text{center}) - f_i(\text{neighbor})|, i = 1, 2$ 
    map disparity to clamped linear indicator function,  $V_c$ 
    compare an existing label  $l$  to its neighbor  $l_n$ 
    if  $l = l_n$  then
      add indicator  $V_c$  to local site energy  $E(l)$ 
    compare a proposed label  $l^+$  to an existing  $l_n$ 
    if  $l^+ = l_n$  then
      add indicator to local site energy  $E(l^+)$ 
  for  $\lambda = l$  and  $l^+$  do
    compare  $\lambda$  to a horizontal, then vertical neighbor  $l_n$ 
    if binary edge segment exists between  $\lambda$  and  $l_n$ 
      and  $\lambda = l_n$  then
        add edge penalty indicator  $V_l$  to  $E(\lambda)$ 
  if  $E(l^+) < E(l) + T$  then
     $l = l^+$ 
```

5. RESULTS

In general the quality of our algorithm is at least comparable to existing methods, but the method incorporates some distinct advantages. For example, the pixel aggregation method requires seed points whose number and locations are not predictable. Split-and-merge algorithms are sensitive to the scale and shape of partition templates. MRF methods, in contrast, begin with random or uniform labels without seed locations. Labelling is not constrained by artificial, regimented partitions that have no resemblance to the underlying image.

We incorporate edges based on a simplified Canny edge detector as an *a priori* feature. Previous work¹ favored a line process using an annealing algorithm much like the region process (a dual). The line process is used to indicate forbidden labelling patterns across region boundaries. We dispensed with this approach because it was found that the line process converges immediately, and its results are not as good as some common edge detectors. Also, our method has the advantage of allowing edges to be simply treated like any other feature. The edge map contributes to overall site energies additively, although the relative contribution can be easily adjusted.

6. CONCLUSION

Although a fast cooling schedule⁷ (exponential) was used, the annealing process is still too slow for real-time video imagery. For the analysis of static images this method does have good potential. A contribution presented herein is to compute and use the edge image as an additional *a priori* feature. Prior work used an annealing process to derive the edge image, which adds substantially to total running time. Here the edge image is treated as just another feature, allowing its contribution to be simply summed. Results show that the MRF approach has the potential for obtaining excellent segmentation. However more research is required in determining optimum parameters of the model, and more experimentation with additional image features in various combinations is desirable.

7. REFERENCES

1. S. Geman and D. Geman. "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images." *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6(6):721-741, 1984.
2. G. E. Hinton and T. J. Sejnowski. "Learning and relearning in boltzmann machines." In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, pages 282-317, MIT Press, Cambridge, 1986.
3. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. "Optimization by simulated annealing." *Science*, 220(4598), 1983.
4. M. I. Miller, B. Roysam, and K. R. Smith. "Mapping rule-based and stochastic constraints to connection architectures." *Proceedings of SPIE*, Cambridge, 1988.
5. D. W. Murray, A. Kasko, H. Buxton. "A parallel approach to the picture restoration algorithm of Geman and Geman on an SIMD." *Image and Vision Computing*, 4:133-142, 1986.
6. F. Spitzer. "Markov random fields and Gibbs ensembles." *Amer. Math. Monthly*, 78:142-154, 1971.
7. H. Tzu. "Fast simulated annealing." In *AIP Conference Proceedings 151. Neural Networks for Computing*, pages 420-245, American Institute of Physics, 1986. Snowbird, UT.