Hierarchical Markov random fields
with stable points

Gregory Budzban and Toney Stirewalt

Southern Illinois University
Department of Mathematics
Carbondale, IL 62901

ABSTRACT

Markov random field techniques for region labeling have become prevalent in image processing research since the seminal work of Geman and Geman in the early 1980's. Their use in actual working systems, however, has been hampered by a number of difficult problems. Perhaps the most intractable of the problems has been the convergence rate of the algorithm. In this paper, we present a technique that introduces stable points in the labeling array of the random field. The stable points are determined by using a simple statistical pixel classifier together with a confidence measure at each pixel. The most confident (top 1%) pixel labels are selected and these labels are used to initiate the evolution of the random field. The stable points introduce pockets of "certainty" in the evolution of the process. The labeling is locally stable and even small numbers of stable points vastly decrease convergence rates of the algorithm.

1. INTRODUCTION

Vision is our most complex and powerful sense. With it, we possess the ability to detect changes in our environment at ranges and sensitivities that the other senses can not match. This capability allows us to manipulate and move within the environment with an irreplaceable ease. The human visual system spans activities from the simplest tasks of finding the boundaries of objects in our field of view to the actual identification of the objects and the formation of procedures to interact with our surroundings. Vision is the link between perception and cognition.

Vision is the fundamental technical problem in robotics. The inability to construct full-scale computer vision systems has severely hampered the development of autonomous machines. At least part of the problem stems from the ad hoc nature of the mathematical techniques used in machine vision. Early successes in solving limited problems hindered the development of an overall mathematical model of vision. Instead of determining algorithms that worked for a wide variety of problems, techniques that were entirely data dependent were found. The "theory" of computer vision consisted of a cookbook of algorithms; each algorithm distinct and designed for a slightly different set of imagery.

It was early in the last decade that this situation began to change. A group of researchers at MIT led by David Marr[1] began to stress the hierarchical nature of the vision problem and spoke of the importance for understanding the effect of the mathematical representation of the image on the classification strategy. At approximately the same time, other investigators began to use spatial stochastic processes, known as Markov random fields (MRF's) as the mathematical representation of an image. The image is viewed as a labeled graph with a certain connectivity between the vertices. The label at each vertex is represented by the value of a random variable associated with that vertex. Using this representation they were able to actually prove theorems showing that the algorithms they proposed converged to an optimal solution.

Having a useful mathematical model of vision is only part of the solution. Severe problems remained however, since the convergence rate for the random field algorithms were badly exponential. A vision system, to be useful, must produce answers quickly enough so that the machine can react in a timely manner. Two facts provide hope that a hierarchical vision system that uses MRF's as the image representation will be able to generate "real-time" solutions. Due to the Markovian nature of the representation, the algorithms are completely spatially parallel. The same instruction set is performed at every location in the image. Thus these algorithms are perfectly adapted for single instruction-multiple data (SIMD) parallel processors[2]. In addition, each level in the vision hierarchy is algorithmically isomorphic. The only change from one level to the next is the size and connectivity of the associated graph.

While end to end vision system have been constructed in the past, their performance has been mediocre. Their performance has been hampered, at least in part, by the lack of a global structure. Each piece of the vision hierarchy was considered separate from the others and completely different algorithm techniques were used on individual levels. Yet the human visual system almost certainly did not evolve separate techniques to solve a problem which is hierarchically isomorphic. The ability of MRF algorithms to model each level of the hierarchy with only insignificant change is a reason to suspect that they more closely match the techniques used in human vision.

While individual levels of the vision hierarchy have been successfully modelled with MRF's, no one has ever attempted to produce a full scale version for one main reason. The convergence rates for the individual levels has been so slow, the problem of combining levels in an end to end system has been considered intractable. This paper will describe an algorithm for the evolution of the MRF that has empirically decreased convergence rates by two orders of magnitude.

2. TECHNICAL DISCUSSION

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.

Another class of random fields was introduced by Dobrushin[3]. 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. 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,

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

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 Spitzer[4] proved that these two seemingly distinct formulations 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.

Much of the current interest in MRF approaches stems from the 1984 article by Geman and Geman[5] 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 called a cooling schedule, that assured convergence.

As indicated earlier, a number of researchers have used stochastic techniques in various stages of the vision hierarchy. In considering edge detection, instead of using an MRF formalism Tan et al.[6] use simulated annealing to minimize a cost functional that evaluates the quality of a given configuration of edges. Yet they freely admit that the functional is heuristic and they make no attempt to fit the algorithm into a hierarchical framework or even indicate how it would interact with other parts of an overall vision system.

The computer vision problem that has received the most attention from researchers using MRF's is the region labelling problem[7,8]. Marques et al. come the closest to the approach proposed by this author. They discuss region labelling as a multi-step hierarchical processes and attempt to reduce the computational load in the MRF algorithm by working on what they refer to as a Gaussian pyramid. The levels of the Gaussian pyramid are formed by convolving the original image Guassian filters with specified variances. In the highest level of the pyramid the variance of the Gaussian is large enough so that the convolution is essentially a low pass filter. The fine details of the image are wiped away and if the MRF algorithm is started at the top of the pyramid the algorithm converges in practice much more quickly due to the smoothness of the image at this level. This result is then used to jump start the algorithm at the next finer level in the pyramid. In each case, the convergence is more rapid than normal since a partial solution has already been acquired from the previous level of the pyramid. No attempt is made to make any of the results mathematically rigorous, however, and once again no other portion of the vision hierarchy is mentioned.

The first attempt to use MRF's for object identification was recently done by Modestino and Zhang[9]. For their work, they consider as their input an image that has already been region labelled. The techniques they use for the process are a simple but natural extension of MRF techniques from lower levels of the vision hierarchy. The region labelled image can be thought of as a colored map where no two adjacent regions have the same label. Just as in the standard mathematical formalism for map coloring problems, each region is considered to be the vertex of a graph with edges to all regions with which it shares a boundary. All other portions of the algorithm remain the same. Again, they do not mention or consider any other portion of the vision hierarchy.

3. ALGORITHM

The overall algorithmic procedure can best be understood with respect to the flow chart below (see Fig.1). At the beginning of the algorithm suite, the image is input into both an edge finding algorithm and a pixel classifier, both deterministic. Deterministic edge finders are usually discrete approximations to directional derivatives. There are a wide variety to chose from and nearly any will do for the purposes of this procedure. A pixel classifier is built by assigning a feature vector to each pixel in the image. The components of the feature vector are usually simple numerical features such as a local mean or variance. Using a training set of imagery, pattern clusters are defined by collecting all the feature vectors associated with a certain region (e.g. forest pixels or road pixels) into a cluster and finding the cluster center and covariance matrix. A pixel is now classified as belonging to the cluster to which it is closest (distance from the cluster center normalized by the covariance matrix).

Next a simple confidence threshold is applied to each edge pixel and pixel classification. In the edge pixel case the threshold will depend on the strength of the edges: those where the magnitude of the derivative is within a certain small percentage of the maximum edge magnitude in the image will be kept. In the pixel classification situation, the confidence will be based on the inverse of the distance to the cluster center. Again only those labels with pass a very stringent threshold (in experiments we have used percentages of 1% or lower) will be kept.

These labels are then used to initiate the evolution of their respective random field. The stable points introduce pockets of "certainty" in the evolution of the process. The labeling is locally stable and even small numbers of stable points vastly decrease convergence rates of the algorithm. As indicated above, in simulations on ten bit infra-red imagery of natural scenes, convergence rates have decreased by two orders of magnitude.

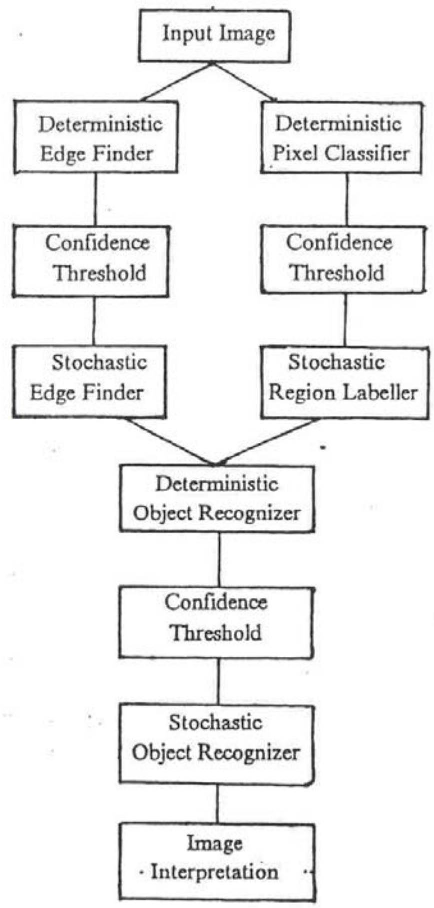

Figure 1. The flow chart for a hierarchical stochastic vision system.

Once the edge and region processes have converge they will be input into a deterministic object recognition algorithm such as the generalized Hough transform[10]. Once again the output of the deterministic algorithm will undergo a confidence threshold and be used to seed the stochastic algorithm. At this point, the graph structure of the MRF is considerable smaller (the vertex set is the number of regions) but the connectivity of the graph will be close to fully connected. In fact, it may be useful to simply make the graph fully connected since when identifying objects occasionally the entire context is required. Finally, the output will be an image interpretation.

4. FUTURE RESEARCH

The introduction of stable points in the evolution process of the random field has been described with respect to a hierarchical computer vision system. A number of interesting theoretical questions are generated with this approach. The resulting Markov chain is not only non-stationary, but also reducible. The labeling configuration on the stable points induces a recurrent, aperiodic class of states. The asymptotic behavior of this type of Markov chain is not well understood. Very few theorems exist containing conditions sufficient for convergence to an invariant measure. Our investigation concerning this topic continues.

5. REFERENCES

[1] Marr, D., *Vision*, W.H. Freeman, San Francisco, (1982).

[2] Budzban, G. and DeCatrel, J., "Markov Random Fields on a SIMD machine for Global Region Labeling", Proceedings of the SPIE, Vol. 1470, pp 175-182, (1991).

[3] Dobrushin, R., "Description of a Random Field by means of Conditional Probabilities", Theory of Probability and its Applications, vol. 13, pp 197-224, (1968).

[4] Spitzer, F., " Markov Random Fields and Gibbs Ensembles", American Math Monthly, vol. 78, pp 142-154, (1971).

[5] Geman, S. and Geman, D., "Stochastic Relaxation, Gibbs Distributions and the Bayesian restoration of Images", IEEE Transactions PAMI, vol. 6, pp 721-741, (1984),

[6] Tan, H., Gelfand, S., and Delp, E., " A Cost Minimization Approach to Edge Detection using Simulated Anealing", IEEE Transactions PAMI, vol. 14, pp 3-18, (1992).

[7] Geman, D., Geman, S., Graffigne, C., and Dong, P., "Boundary Detection by Constrained Optimization", IEEE Transactions PAMI, vol. 12, pp 609-628, (1990).

[8] Marques, F., Cummillera, J., and Gasull, A., "Hierarchical Segmentation Using Compound Gauss-Markov Random Fields", Proceedings IEEE ASSP International Conference, vol. 3, pp 53-56, (1992).

[9] Modestino, J., and Zhang, J., " A Markov Random Field approach to Image Interpretation", IEEE Transactions PAMI, vol.14, pp 606-615, (1992).

[10] Ballard,D., and Brown, C., *Computer Vision*, Prentice-Hall, Englewood Clifs, New Jersey, (1982).