Relaxation Labeling of Markov Random Fields

Stan Z. Li, Han Wang

School of EEE
Nanyang Technological University
Singapore 2263

Maria Petrou

Department of EEE
University of Surrey
Guilford, Surrey GU2 5XH, UK

Abstract

Using Markov random field (MRF) theory, a variety of computer vision problems can be modeled in terms of optimization based on the maximum a posteriori (MAP) criterion. The MAP configuration minimizes the energy of a posterior (Gibbs) distribution. When the label set is discrete, the minimization is combinatorial. This paper proposes to use the continuous relaxation labeling (RL) method for the minimization. The RL converts the original NP complete problem into one of polynomial complexity. Annealing may be combined into the RL process to improve the quality (globalness) of RL solutions. Performance comparison among four different RL algorithms is given.

1 Introduction

Since 1980's, there has been considerable interest in image and vision modeling using Markov random field (MRF) theory [3]. MRF theory provides us with a tool for modeling a vision problem within the established Bayes framework. In MRF based Bayes modeling, a problem is posed as one of labeling. When the interaction between the labels are limited within a neighborhood, the labeling configuration can be regarded as an MRF. The joint posterior probability of an MRF obeys a Gibbs distribution. The corresponding (posterior) energy function is then used to measure the cost of a solution. The MAP solution minimizes the posterior energy.

A labeling problem is specified in terms of a set of sites and a set of labels. Let

$$\mathcal{S} = \{1, \ldots, m\} \tag{1}$$

be a set of m discrete sites. In this paper, we considered a set of M discrete labels

$$\mathcal{L} = \{1, \ldots, M\} \tag{2}$$

Labeling is to assign a label from $\mathcal{L}$ to each of the sites in $\mathcal{S}$. Define a neighborhood system for $\mathcal{S}$

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\} \tag{3}$$

where $\mathcal{N}_i$ is the collection of sites neighboring i. A clique c for $(\mathcal{S}, \mathcal{N})$ is a subset of $\mathcal{S}$ consisting either of a single site or of several neighboring sites.

Let $F = \{F_1, \ldots, F_m\}$ be a family of random variables defined on $\mathcal{S}$, in which each random variable F_i assumes a value in $\mathcal{L}$. F is an MRF on $\mathcal{S}$ with respect to $\mathcal{N}$ if and only if the probability distribution $P(F = f)$ of the configurations is a Gibbs distribution with respect to $\mathcal{N}$. A Gibbs distribution of the configurations f with respect to $\mathcal{N}$ is of the following form

$$P(f) = Z^{-1} \times e^{-U(f)/T} \tag{4}$$

In the above, Z is a normalizing constant, T is a global control parameter called the temperature and $U(f)$ is the prior energy. When cliques containing at most two sites are considered, the energy has the form

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) = \sum_{\{i\} \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i,j\} \in \mathcal{C}_2} V_2(f_i, f_j) \tag{5}$$

We are interested in maximizing posterior probabilities which are also of the Gibbs form:

$$P(F = f \mid r) = Z_E^{-1} \times e^{-E(f)} \tag{6}$$

where r is the observation and the energy, $E(f) = \frac{1}{T}U(f) + U(r \mid f)$, is a sum of the prior energy $U(f)$ and the likelihood energy $U(r \mid f)$. The energy can be written

$$E(f) = \sum_{i \in \mathcal{S}} V_1(f_i \mid r_1(i)) + \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} V_2(f_i, f_j \mid r_2(i,j)) \tag{7}$$

where $r_1(i)$ and $r_2(i,j)$ are unary and binary components of the observation, and

$$V_1(f_i \mid r_1(i)) = V_1(f_i)/T + V_1(r_1(i) \mid f_i) \tag{8}$$

and

$$V_2(f_i, f_j \mid r_2(i,j)) = V_2(f_2, f_j)/T + V_2(r_2(i,j) \mid f_i, f_j) \tag{9}$$

are the posterior potentials. The MAP solution is equivalently found by minimizing the posterior energy

$$f^* = \arg\min_{f \in \mathbb{F}} E(f) \tag{10}$$

where $\mathbb{F} = \mathcal{L}^m$ is the admissible configuration space for f.

When $\mathcal{L}$ consists of discrete labels, the minimization of the posterior energy is combinatorial and thus NP complete. Moreover, there can be many local minima in the solution space which affect solution quality. Stochastic annealing algorithms such as simulated annealing [11, 8] are often used for non-convex and combinatorial optimization are used for minimizing MRF energies. They are proven to find the global solution with probability approaching one. However, such algorithms are well-known to be expensive and inefficient. The efficiency can be improved by using a deterministic approximation, such as graduated non-convexity [1], mean field theory approximation [6] and the highest confidence first (HCF) algorithm [4]. But these are for low level MRF models defined on a regular lattice like image grid and do not yet cater for labeling of arbitrary sites.

In this section, we use the relaxation labeling (RL) method as a deterministic means for solving the combinatorial minimization problem. RL [18] is a class of iterative algorithms which assign a label from a discrete label set to each of the sites to satisfy certain constraints. In RL, constraints are propagated via a compatibility function and the ambiguity of labeling is reduced as the iteration continues. In Section 2, the original combinatorial minimization problem is converted into one of continuous RL having polynomial complexity. An annealing RL (ARL) algorithm is described for improving the globalness. The ARL has been used for object recognition [13]. In Section 3, four different RL algorithms are compared in terms of the globalness of solutions and convergence rate.

2 Combinatorial Energy Minimization using RL

Indeed RL can be regared as a computational mechanism for solving combinatorial minimization problems. Formally, Faugeras and Berthod [5] define a class of global criteria in terms of transition probabilities . Haralick [9] illustrates RL as minimizing expected loss. Hummel and Zucker [10] show that, computationally, finding consistent labeling is equivalent to solving a variational inequality. Kittler *et al.* [12] give a theoretical explaination of probabilistic relaxation using the Bayesian framework.

2.1 Converting MRF Configuration to Labeling Assignment

To apply continuous RL, we use a labeling assignment to denote an MRF labeling configuration. In continuous RL, the labeling assignment is defined by[1]

$$f = \{f_{i,I} \in [0,1] \mid \sum_{I \in \mathcal{L}} f_{i,I} = 1, i \in \mathcal{S}, I \in \mathcal{L}\} \tag{11}$$

where $\sum_{I \in \mathcal{L}} f_{i,I} = 1,\ i \in \mathcal{S}$ is the consistency constraint. The real value $f_{i,I} \in [0,1]$ reflects the strength with which i is labeled I. Let $\bar{f}_i = (f_{i,1}, \ldots, f_{i,M})$ be the (fuzzy) labeling state of site i. The consistency constraint confines each $\bar{f}_i$ to lie in a hyperplane in the M dimensional real space R^M. This defines the admissible space for the assignment f

$$\mathbb{F} = \{f \mid f_{i,I} \in [0,1], \sum_{I \in \mathcal{L}} f_{i,I} = 1, i \in \mathcal{S}, I \in \mathcal{L}\} \tag{12}$$

The final solution f^* is subject to an additional constraint, the unambiguity constraint

$$f^*_{i,I} = \{0,1\} \qquad \forall i, I \tag{13}$$

with $f^*_{i,I} = 1$ meaning that i is unambiguously mapped to I by f^*. This defines the final admissible space for f^*

$$\mathbb{F}^* = \{f^* \mid f_{i,I} \in \{0,1\}, \sum_{I \in \mathcal{L}} f_{i,I} = 1, i \in \mathcal{S}, I \in \mathcal{L}\} \tag{14}$$

The $\mathbb{F}^*$ is the set of "corners" of the $\mathbb{F}$. Points in the continuous space $\mathbb{F}$ provides routes to the corners.

2.2 Converting Potentials to Compatibilities

Now we convert the posterior potential functions in the MRF formulation to compatibility functions in the RL. The posterior potentials $V_1(f_i \mid r_1(i))$ measure the costs incurred by each individual labels $i \to f_i$ and $V_2(f_i, f_j \mid r_2(i,j))$ measure the costs incurred by each pair of labels $i \to f_i$ and $j \to f_j$. The compatibilities measures the "gains" instead. The compatibility

[1] To keep notations tractable, the same notation f is used to denote both a label configuration and the corresponding labeling assignment.

functions are related to the potentials by the following relationships

$$G_1(I \mid i) = C - V_1(f_i \mid r_1(i)) \tag{15}$$

and

$$G_2(I, J \mid i, j) = C - V_2(f_i, f_j \mid r_2(i, j)) \tag{16}$$

where $I = f_i$ and $J = f_j$ and C is a constant. Recall that there is a temperature parameter T in the prior distribution. Therefore the posterior potentials $V_1(f_i \mid r_1(i))$ and $V_2(f_i, f_j \mid r_2(i, j))$ are parameterized by T and so are the compatibility functions $G_1(I \mid i)$ and $G_2(I, J \mid i, j)$.

Using the notation of the labeling assignment and the compatibility functions, we convert minimizing the MRF energy $E(f)$ to maximizing a gain function $G(f)$. The gain is defined by

$$\begin{aligned} G(f \mid T) &= \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} G_1(I \mid i)\ f_{i,I} + \\ & \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} \sum_{j \in \mathcal{S}} \sum_{J \in \mathcal{L}} G_2(I, J \mid i, j)\ f_{i,I}\ f_{j,J} \end{aligned} \tag{17}$$

The temperature T is now expressed explicitly in the gain notation. It can be used as the control parameter for annealing. The target gain function is defined under $T = 1$.

To reduce storage, in experiments, we use only one byte (8 bits) to represent the compatibility coefficients. A positive coefficient is truncated to an integer between 0 and 255 while negative values are truncated to zeros. In this case, we set $C = 255$.

Now the minimization in (10) is converted to the following constrained maximization

$$f^* = \max_{f \in \mathbb{F}^*} G(f \mid T = 1) \tag{18}$$

The maximization is constrained because any labeling assignment f must satisfy the consistency and the unambiguity constraints. The above is still a combinatorial problem because all $f_{i,I}$ in $f \in \mathbb{F}^*$ assumes a value of 1 or 0. In continuous RL, it is further converted to a real problem in which the maximization is done via the space $\mathbb{F}$

$$f^* = \max_{f \in \mathbb{F}} G(f \mid T = 1) \tag{19}$$

Using appropriate RL algorithms, the final f^* will be at one of the corners of $\mathbb{F}$, *i.e.* $f^* \in \mathbb{F}^*$ [10].

2.3 Iterative RL Procedures

RL algorithms perform constrained maximization based on the gradient information $\nabla G(f^)$. Introduce a time parameter into f such that $f = f^{(t)}$ represents the state of labeling at time t. The assignment is initially set to an $f^{(0)} \in \mathbb{F}$. At time t, the gradient function

$$q^{(t)} = \nabla G(f^{(t)}) \tag{20}$$

is computed. Then $f^{(t+1)}$ is obtained using $f^{(t)}$ and $G(f^{(t)})$. RL algorithms differ in the way they update $f^{(t)}$ into $f^{(t+1)}$.

Rosenfeld *et al.* propose to use the following fixed point iteration

$$f_{i,I}^{(t+1)} \leftarrow \frac{f_{i,I}^{(t)}(1 + q_{i,I}^{(t)})}{\sum_I f_{i,I}^{(t)}(1 + q_{i,I}^{(t)})} \tag{21}$$

where $|q_{i,I}^{(t)}| \leq 1$ is assumed. Obviously, the consistency is maintained by the above because $f_i^{(t+1)}$ is a normalized vector. The above iteration equation is only one of many choices. When the compatibility coefficients are non-negative such that $q_{i,I}^{(t)} \geq 0$, the following can be used

$$f_{i,I}^{(t+1)} \leftarrow \frac{f_{i,I}^{(t)} q_{i,I}^{(t)}}{\sum_I f_{i,I}^{(t)} q_{i,I}^{(t)}} \tag{22}$$

In the continuous RL algorithms by Faugeras and Berthod [5] and Hummel and Zucker [10], the consistency are maintained using a technique called gradient projection (GP). The use the following updating rule

$$f^{(t+1)} \leftarrow f^{(t)} + \Psi(f^{(t)}, q^{(t)}) \tag{23}$$

where the $\Psi(f^{(t)}, q^{(t)})$ is a GP operator. The operator returns an updating vector (direction plus magnitude) which ensures that the updated vector $f^{(t+1)}$ lies within the space $\mathbb{F}$. The convergence to a point in $\mathbb{F}^*$ is generally guaranteed by the Hummel-Zucker algorithm [10].

Peterson and Soderberg use the mean field approach to minimize energy function [16]. They derive the following fixed point iteration for the updating

$$f_{i,I}^{(t+1)} \leftarrow \frac{e^{q_{i,I}^{(t)} \tau}}{\sum_I e^{q_{i,I}^{(t)}/\tau}} \tag{24}$$

where τ is an annealing parameter which should be decreased to 0^+. The consistency is maintained by the above. The unambiguity of f is forced when $\tau \to 0^+$.

In our experiments, τ is decreased by using $\tau^{(t+1)} \leftarrow 0.95\tau^{(t)}$. The convergence of RL can be judged based by a quantity called "saturation" of f which is defined as $S = \frac{1}{m}\sum_I f_{i,I}^2$. An RL procedure is considered converged if $S > 0.99$.

2.4 Annealing RL

RL algorithms are local optimizers by nature. They do not necessarily give the global optimum. Given an RL algorithm and the data, two factors affect the solution: the initial assignment and the compatibility function [18]. However, it is shown that for certain choices of compatibility function, an RL algorithm can have a *single* convergence point regardless of the initial assignment [2, 14].

In the traditional RL [18, 15], compatibility functions are defined without referring to the observation (and thus should be interpreted as being encoding *a priori* constraints); constraints due to the observation is encoded in the initial assignment. Therefore, traditional RL solutions rely much on the initialization and convergence into one point regardless of the initialization is considered undesirable. In contrast, the posterior MRF energy has already contained the both sources of constraints, which is a progress over the traditional RL formulation, and therefore the global energy minimum is the desired solution [8].

To improve the globalness of solutions, we combine annealing into relaxation labeling, yielding annealing RL (ARL). In ARL, the annealing is performed by cooling down the temperature T during RL process. This improves the quality of MRF solutions. The ARL has been used to obtain MRF solutions for MRF object recognition. In the annealing schedule, T is initially set to a very high value $T^{(0)} \rightarrow \infty$ and gradually lowered ($T^{(t)} < T^{(t-1)}$) to $T^{(\infty)} = 1$. The RL solutions f^* are tracked from high T to low T. The maximum f^* obtained at $T^{(t-1)}$ is used as the initial value for the next new phase of RL iterations at $T^{(t)}$. By this, we can obtain an approximation of the global solution at the target T. The ARL is described in fig.1.

Annealing Relaxation Labeling of MRFs
Begin Algorithm
(1) set $t \leftarrow 0$, $T \leftarrow 100$; randomly set $f \in \mathbb{F}$;
(2) do {
(3) lower temperature $T \leftarrow \text{lower}(T)$;
(4) do { /* RL under T */
(5) compute gradient $q = \nabla G(f \mid T)$;
(6) compute gradient projection $\delta f = \Psi(f, q)$;
(7) update labeling assignment $f \leftarrow f + \delta f$;
(8) } until ($\|\delta f\| < \epsilon$); /* converged under T */
(9) $t \leftarrow t + 1$;
(10) } until ($T = 1$) /* reached the target value */
End Algorithm

Figure 1: The ARL algorithm.

3 Experimental Results

The experiments are based on the setting of matching simulated graphs. First, we generate a weighted graph at random. Then we randomly deviate the weights, remove some nodes and add some nodes, generating another weighted graph. An MRF model for object recognition [13] is used to derive the posterior energy. The energy is minimized using four different RL algorithms: simulated anneling [11, 8] with the annealing schedule set as $T \leftarrow 0.999T$, the classical continuous RL [18], the RL using gradient projection [10] and the mean field theory RL [16]. Fig.2 gives quantitative comparison of different algorithms in solution quality measured by the gain and convergence rate measured by the iteration numbers. All algorithms are tested with the same energies derived from the same pairs of weighted graphs; furthermore, the three deterministic RL algorithms are initialized in the same way. The statistics are obtained by averaging 100 results.

The plot on the right shows that the iteration number required by the SA algorithm is about 10^4 times the rest, the MFT algorithm of Peterson-Soderberg has the fastest convergence and the Hummel-Zucker algorithm has a very stable convergence rate. The plot on the left shows that with such expenses, the SA does not necessarily yield better results the deterministic RLs though in theory, the solution quality can be improved using a schedule which is slow enough. The plot also demonstrates that when noise is low, the deterministic algorithms performs better; whereas when noise is high, simulated annealing gives average results of better quality. This conforms the observation made by [7] about the two classes of algorithms. The reason that when noise is higher, the problem of multiple local optima becomes more significant and simulated annealing is better in tackling this problem.

In comparison to an earlier evaluation work [17] in which the quality of a solution is measured by

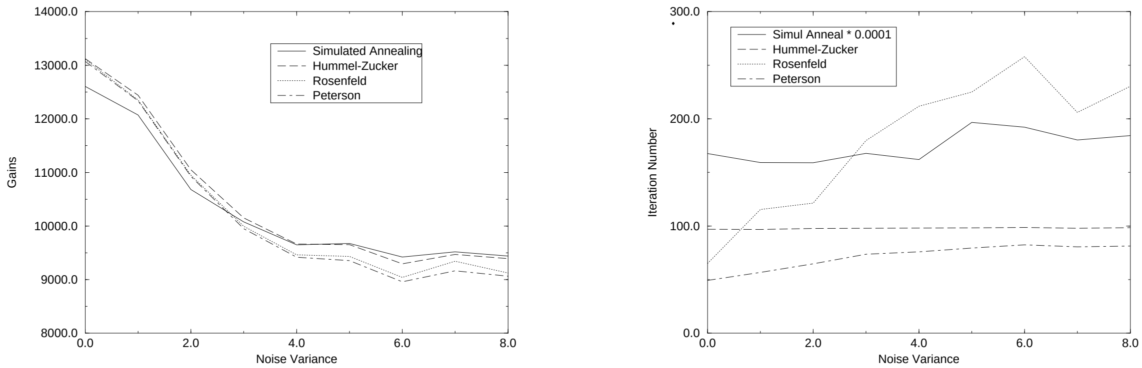

Figure 2: Comparison of gain and convergence. Left: Gain of solutions obtained by four algorithms (the higher the better). Right: Number of iterations for the algorithms to converge (the lower the better); the plotted number for the SA is scaled by a factor of 10^{-4}.

the goodness of interpretation, the quality in this work is measured by the quantitative remaining gain. The evaluation there is not only related to algorithms themselves but also to how the problem is formulated (*e.g.* definition of compatibilities), mixing the two issues. Here, we regard RL as just a mechanism of minimization rather than one of interpretation; therefore we concern only how well an algorithm can minimize an energy regardless of the problem domain. On the other hand, each of the measures in Fig.2 are obtained by averaging many test results. In [17], conclusions are drawn based on a few particular cases. The method used here should be more reliable and objective.

4 Conclusion

The motivation of this work has been to minimize MRF energy in a discrete configuration space in an economical and efficient way. We have proposed to use the relaxation labeling (RL) method for this purpose. By this means, the original NP complete problem is converted into a polynomial one. The deterministic RL is preferred in that it is much cheaper than stochastic methods such as simulated annealing, yet it gives results of comparable quality.

We have also presented a method for improving the quality (globalness) of the RL solution by combining annealing into the RL process. We found that the annealing improves the quality of results (the gain of solutions) in these object recognition experiments. Without it, the RL tends to give less favorable results.

References

[1] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.

[2] H. I. Bozma and J. S. Duncan. "Admissibility of constraint functions in relaxation labeling". In *Proceedings of Second International Conference on Computer Vision*, pages 328–332, Florida, December 1988.

[3] R. Chellappa and A. Jain, editors. *Markov Random Fields: Theory and Applications*. Academic Press, 1993.

[4] P. B. Chou and C. M. Brown. "The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4:185–1.810, 1990.

[5] O. D. Faugeras and M. Berthod. "Improving consistency and reducing ambiguity in stochastic labeling: An optimization approach". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3:412–423, April 1981.

[6] D. Geiger and F. Girosi. "Parallel and deterministic algorithms from mrf's: surface reconstruction". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5), May 1991.

[7] D. Geman, G. Geman, C. Graffigne, and P. Dong. "Boundary detection by constrained optimization". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-12(7):609–628, July 1990.

[8] G. Geman and D. Geman. "Stochastic relaxation, gibbs distribution and bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[9] R. M. Haralick. "Decision making in context". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(4):417–428, July 1983.

[10] R. A. Hummel and S. W. Zucker. "On the foundations of relaxation labeling process". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(3):267–1.886, May 1983.

[11] S. Kirkpatrick, C. D. Gellatt, and M. P. Vecchi. "Optimization by simulated annealing". *Science*, 220:671–680, 1983.

[12] J. Kittler, W. J. Christmas, and M. Petrou. "Probabilistic relaxation for matching problem in computer vision". In *Proceedings of Forth International Conference on Computer Vision*, pages 666–673, Gemany, May 1993.

[13] S. Z. Li. "A Markov random field model for object matching under contextual constraints". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, Washington, June 1994.

[14] D. P. O'Leary and S. Peleg. "Analysis of relaxation processes: the two-node two label case". *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13(4):618–623, 1983.

[15] S. Peleg and R. A. Rosenfeld. "Determining compatibility coefficients for curve enhancement relaxation processes". *IEEE Transactions on Systems, Man and Cybernetics*, 8, 1978.

[16] C. Peterson and B. Soderberg. "A new method for mapping optimization problems onto neural networks". *International Journal of Neural Systems*, 1(1):3–1.82, 1989.

[17] K. E. Price. "Relaxation matching techniques – A comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(5), September 1985.

[18] A. Rosenfeld, R. Hummel, and S. Zucker. "Scene labeling by relaxation operations". *IEEE Transactions on Systems, Man and Cybernetics*, 6:420–433, June 1976.