Journal of Mathematical Imaging and Vision, 7, 1–162 (1997)

Energy Minimization and Relaxation Labeling*

STAN Z. LI, HAN WANG, KAP LUK CHAN
School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore 639798

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

Received . Revised .

Abstract. Recently, there has been increasing interest in Markov random field (MRF) modeling for solving a variety of computer vision problems formulated in terms of the *maximum a posteriori* (MAP) probability. When the label set is discrete, such as in image segmentation and matching, the minimization is combinatorial. The objective of this paper is twofold: Firstly, we propose to use the continuous relaxation labeling (RL) as an alternative approach for the minimization. The motivation is that it provides a good compromise between the solution quality and the computational cost. We show how the original combinatorial optimization can be converted into a form suitable for continuous RL. Secondly, we compare various minimization algorithms, namely, the RL algorithms proposed by Rosenfeld *et al.* and by Hummel and Zucker, the mean field annealing of Peterson and Soderberg, simulated annealing of Kirkpatrick, the iterative conditional modes (ICM) of Besag and an annealing version of ICM proposed in this paper. The comparisons are in terms of the minimized energy value (*i.e.* the solution quality), the required number of iterations (*i.e.* the computational cost), and also the dependence of each algorithm on heuristics.

Keywords: Contextual constraints, constrained optimization, Markov random field (MRF), maximum *a posteriori* (MAP), relaxation labeling

1. Introduction

Since its beginning, computer vision research has been evolving from heuristic design of algorithms to systematic investigation of approaches. In this process, researchers have realized (1) that due to uncertainties, the solution to a vision problem cannot be exact but only in a sense of optimal satisfaction of a set of constraints and (2) that *contextual constraints* are ultimately necessary for the understanding of visual information in images. This raises the questions of how to define the optimal solution under contextual constraints and how to find it.

Markov random field (MRF) theory, a branch of probability theory, provides a foundation for the characterization of spatial, contextual constraints and the derivation of the probability distributions of interacting features. In conjunction with methods from decision and estimation, MRF theory provides a systematic approach for modeling vision and image analysis problems based on established optimality criteria such as *maximum a posteriori* (MAP) probability [16]. This MAP-MRF framework enables us to develop vision algorithms for a variety of problems systematically using sound principles rather than *ad hoc* heuristics.

An optimization procedure is required to find the MAP configuration. It is combinatorial when the label set is discrete. Traditionally, the combinatorial optimization is performed by using an algorithm operating on the corresponding discrete solution space, such as simulated annealing [11], [8] or iterative conditional modes (ICM) [2].

* This work was supported by NTU project RG 43/95.

In this paper, we consider the continuous relaxation labeling (RL) method as a deterministic mechanism for the combinatorial optimization. RL, proposed by Rosenfeld *et al.* [30], 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. The ambiguity of labeling is reduced as the iteration continues. It has been widely used in image and vision community for image enhancement, edge detection, object matching and so on.

RL can be regarded as a computational mechanism for solving combinatorial minimization problems. Formally, Faugeras and Berthod [6] define a class of global criteria in terms of transition probabilities. Peleg [25] interprets RL in terms of Bayes analysis. Haralick [9] illustrates RL as minimizing the expected loss from a Bayes viewpoint. Hummel and Zucker [10] formulate the probabilistic relaxation approach as a global optimization problem in which a so-called average local consistency is defined and maximized. They show that, computationally, finding consistent labeling is equivalent to solving a variational inequality. However, RL is not readily applicable to the combinatorial optimization and a proper conversion is necessary.

We will show how a combinatorial minimization problem can be transformed into a form suitable for continuous RL. More specifically, we convert the original optimization over a discrete space into a problem of constrained real optimization over a continuous space. The converted form is then solved by using a continuous RL method such as Rosenfeld *et al.* [30], Hummel and Zucker [10], or Peterson and Soderberg [27].

We will then compare various algorithms in terms of the minimized energy value (*i.e.* the solution quality), the required number of iterations (*i.e.* the computational cost), and also the dependence of each algorithm on heuristics. The compared algorithms are the above mentioned the RL algorithms of [30], [10] and [27], simulated annealing [11], ICM [2], and an improved, annealing version of ICM for the purpose of matching. The results of the comparisons are in favor of the Hummel-Zucker algorithm.

The paper is organized as follows: Section 2 formulates the conversion from the combinatorial optimization to a form suitable for continuous RL. Several RL iteration schemes are described and related issues discussed. An MAP-MRF matching formulation is described in Section 3 in the notion of relational graphs. It is then used as the testbed in the experiment comparisons in Section 4. Conclusion is given in Section 5.

2. Energy Minimization and Relaxation Labeling

2.1. *MAP-MRF Labeling*

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.

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

be a set of M discrete labels. Labeling is to assign a label from $\mathcal{L}$ to each of the sites in $\mathcal{S}$.

An MRF is always defined with respect to a neighborhood system. 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. To define an MRF in terms of a Gibbs distribution, we need to introduce cliques. 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 called the energy. When cliques containing at most two sites are considered, the energy has the form

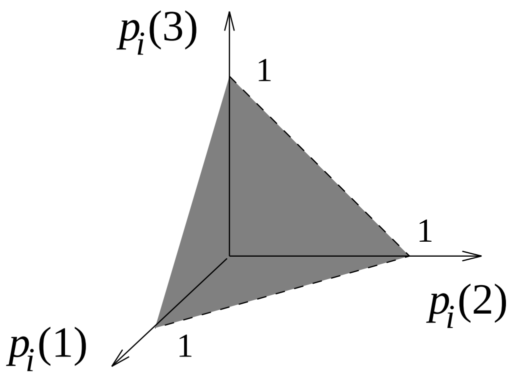

Fig. 1. The labeling space for p_i with three labels. The feasible space is shown as the shaded area (simplex) and the unambiguous space consists of the three corner points.

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

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are the set of single-site and the set of pair-site cliques, respectively.

In the MAP-MRF framework [8], a vision problem, such as image restoration and object matching, can be formulated as maximizing the posterior probability

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

where d is the observation and the energy, $E(f) = \frac{1}{T}U(f)+U(d \mid f)$, is a sum of the prior energy $U(f)$ and the likelihood energy $U(d \mid f)$. A special example of MAP-MRF model for object matching will be described in Section 3. The MAP solution is equivalently found by minimizing the posterior energy

$$f^* = \arg \min_{f \in \mathbb{F}} E(f) \quad (7)$$

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.

2.2. Continuous Relaxation Labeling

By the notion of continuous RL, the combinatorial minimization is converted into a real minimization subject to linear constraints, and the combinatorial minimization is performed by constrained real minimization.

In continuous RL, the assignment of a site is represented as

$$p_i = [p_i(I) \mid I \in \mathcal{L}] \quad (8)$$

where p_i is subject to the following *feasibility* constraints

$$\begin{cases} \sum_{I \in \mathcal{L}} p_i(I) = 1 & \forall i \\ p_i(I) \geq 0 & \forall i, I \end{cases} \quad (9)$$

This is interpreted as a fuzzy model [30]. The real value $p_i(I) \in [0,1]$ reflects the strength with which i is assigned label I. Each p_i lies in a hyperplane in the positive quadrant portion (simplex) of the multi-dimensional real space $\mathbb{R}^{\mathcal{L}}$ as shown in the shaded area of Fig.1. The feasible labeling assignment space is

$$\mathbb{P}_i = \{p_i(I) > 0 \mid I \in \mathcal{L}, \sum_{I \in \mathcal{L}} p_i(I) = 1\} \quad (10)$$

The $p = \{p_i \mid i \in \mathcal{S}\}$ is called a *labeling assignment*. The feasible space for the labeling assignment is the following product

$$\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2 \cdots \times \mathbb{P}_m \quad (11)$$

The final solution p^* must be *unambiguous*

$$p_i^*(I) = \{0, 1\} \qquad \forall i, I \quad (12)$$

with $p_i^*(I) = 1$ meaning that i is unambiguously labeled I. This gives the unambiguous spaces for p_i as

$$\mathbb{P}_i^* = \{p_i(I) \in \{0,1\} \mid I \in \mathcal{L}, \sum_{I \in \mathcal{L}} p_i(I) = 1\} \quad (13)$$

So every p_i in $\mathbb{P}_i^*$ is one of the vectors $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$ and $(0, 0, \ldots, 1)$ corresponding to the "corners" of the simplex $\mathbb{P}_i$ (*cf.* Fig.1). The unambiguous labeling assignment space is

$$\mathbb{P}^* = \mathbb{P}_1^* \times \mathbb{P}_2^* \cdots \times \mathbb{P}_m^* \quad (14)$$

which consists of the corners of the space $\mathbb{P}$. Points in the continuous space $\mathbb{P}$ provide routes to the corners in the discrete space $\mathbb{P}^*$. The unambiguity may be satisfied automatically upon convergence of some iterative algorithms or otherwise has to be enforced for example using a maximum selection (winner-take-all) operation. An unambiguous labeling assignment is related to the corresponding discrete labeling by

$$f_i = I \qquad if \quad p_i^*(I) = 1 \tag{15}$$

2.3. Maximization Formulation

Although RL can be used for both minimization and maximization, it is sometimes algorithmically more suitable for maximization. Therefore, in practice, we convert the minimization of the MRF energy into the maximization of a corresponding gain function. The gain is the sum of compatibility functions.

The RL compatibility functions can be defined based on the Gibbs clique potential functions as follows: The unary compatibility function is defined by

$$r_i(I) = Const_1 - V_1(I \mid d) \tag{16}$$

where $V_1(I \mid d)$ is the function of potentials incurred by single-site cliques. The binary compatibility function is defined by

$$r_{i,i'}(I, I') = Const_2 - V_2(I, I' \mid d) \tag{17}$$

where $V_2(I, I' \mid d)$ is the function of potentials incurred by pair-site cliques. The constants $Const_1$ and $Const_2$ are chosen so that all the compatibility functions are non-negative.

The gain with the labeling assignment representation can now be written as

$$G(f) = \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} r_i(I)\ p_i(I) + \\ \sum_{i \in \mathcal{S}} \sum_{I \in \mathcal{L}} \sum_{i' \in \mathcal{S}} \sum_{I' \in \mathcal{L}} r_{i,i'}(I, I')\ p_i(I)\ p_{i'}(I') \tag{18}$$

which is to be *maximized.* The above is the standard form of objective functions in RL formulations. Obviously, there is one-one correspondence between the maxima of $G(f)$ and the minima of $E(f)$ because of the relationship $G(f) \propto Const - E(f)$ where $Const$ is some constant.

2.4. Iterative Updating Equations

In maximizing $G(f)$ under the feasibility and unambiguity constraints, an RL algorithm iteratively updates p based on the gradient $q = -\nabla E(f)$. The components of $q = \{q_i(I) \mid \forall(i, I)\}$ are given by

$$q_i(I) = \frac{\partial G}{\partial p_i(I)} = \\ r_i(I) + 2 \sum_{i' \in \mathcal{S}} \sum_{I' \in \mathcal{L}} r_{i,i'}(I, I')\ p_{i'}(I') \tag{19}$$

with symmetric $r_{i,i'}(I, I')$. Let $p = p^{(t)}$ represent the labeling state at time t. The assignment is initially set to an $p^{(0)} \in \mathbb{P}$. The current $p^{(t)}$ is updated into the new state $p^{(t+1)}$ using $p^{(t)}$ and $\nabla G(p^{(t)})$. RL algorithms differ from one another in the way $p^{(t)}$ is updated into $p^{(t+1)}$. Various update equations are described in the following.

Rosenfeld *et al.* [30] propose to use the following fixed point iteration

$$p_i^{(t+1)}(I) \leftarrow \frac{p_i^{(t)}(I)[1 + q_i^{(t)}(I)]}{\sum_J p_i^{(t)}(J)[1 + q_i^{(t)}(J)]} \tag{20}$$

where $-1 \le q_i^{(t)}(I) \le 1$ is assumed. Obviously, the feasibility constraints (9) are satisfied by the normalization in the formula. The above iteration equation is only one of many choices. For example, when the compatibility coefficients are non-negative, we have $q_i^{(t)}(I) \ge 0$; in this case, the following updating scheme also maintains the feasibility while performing gradient ascent

$$p_i^{(t+1)}(I) \leftarrow \frac{p_i^{(t)}(I) q_i^{(t)}(I)}{\sum_J p_i^{(t)}(J) q_i^{(t)}(J)} \tag{21}$$

In the continuous RL algorithms by [6] and [10], the feasibility is maintained using a technique called *gradient projection* (GP) [29]. GP is an extended steepest descent procedure for solving nonlinear programming problems under linear constraints. It finds a feasible direction u, from the gradient q and the current $p^{(t)}$, for the following updating scheme

$$p^{(t+1)} \leftarrow p^{(t)} + \mu u \tag{22}$$

where $\mu > 0$ is a scalar. The vector $u = \frac{1}{\mu}[p^{(t+1)} - p^{(t)}]$ must satisfy the following conditions

$$\begin{aligned}\sum_{I \in \mathcal{L}} u_i(I) &= \frac{1}{\mu} \sum_{I \in \mathcal{L}} [p_i^{(t+1)}(I) - p_i^{(t)}(I)] \\ &= \frac{1}{\mu}(1-1) = 0\end{aligned} \tag{23}$$

in order to maintain the feasibility constraints. If $p_i^{(t)}(I) = 0$, then there must be $u_i(I) \geq 0$ so that $p_i^{(t+1)}(I)$ would not be negative. For convenience, a normalization of u may be imposed so that $\|u\| = 1$ where $\|\cdot\|$ is the Euclidean norm. Then, the feasible space for u is defined by

$$\begin{aligned}&\mathbb{U} = \{u \in \mathbb{R}^{\mathcal{S} \times \mathcal{L}} \mid \\ &\sum_{I \in \mathcal{L}} u_i(I) = 0, \|u\| = 1, u_i(I) \geq 0 \text{ whenever } p_i(I) = 0\}\end{aligned} \tag{24}$$

(note that $u_i(I)$ can be negative when $p_i(I) > 0$). With u confined in the above space, $\sum_I p_i^{(t+1)}(I) = 1$ are satisfied. The conditions $0 \leq p^{(t+1)} \leq 1$ are met by choosing a small enough μ for which the following inequalities are satisfied

$$0 \leq p_i^{(t+1)}(I) = p_i^{(t)}(I) + \mu u_i(I) \leq 1 \qquad \forall i, I \tag{25}$$

A detailed implementation for the GP operator used in [10] is described in [22]. A discussion on parallel distributed implementation of the Hummel-Zucker algorithm in multiprocessor architecture environment is given in [20].

Mean field theory provides another approach for solving combinatorial optimization problems using real computation. Peterson and Soderberg [27] derive an alternative iteration equation, which is consistent with the continuous RL representation, as follows.

Let $f = \{f_i(I) \in \{0,1\} \mid i \in \mathcal{S}, I \in \mathcal{L}\}$ be an unambiguous labeling assignment in $\mathbb{P}^*$ defined in (14); each f_i takes the value of one of the vectors $(1,0,\dots,0)$, $(0,1,\dots,0)$, and $(0,0,\dots,1)$. Denote the energy with the RL representation still by $E(f)$. The Gibbs distribution of the energy under temperature $T > 0$ is given as

$$P_T(f) = Z^{-1} e^{-\frac{1}{T}E(f)} \tag{26}$$

where the partition function

$$Z = \sum_{f \in \mathbb{P}^*} e^{-\frac{1}{T}E(f)} \tag{27}$$

sums over $\mathbb{P}^*$. The statistical mean of the distribution is defined as

$$\langle f \rangle_T = \sum_f f P_T(f) = \sum_f f Z^{-1} e^{-\frac{1}{T}E(f)} \tag{28}$$

As the temperature approaches zero, the mean of the Gibbs distribution approaches its mode or the global energy minimum f^*

$$\lim_{T \to 0} \langle f \rangle_T = \lim_{T \to 0} \sum_f f P_T(f) = f^* \tag{29}$$

This suggests that instead of minimizing $E(f)$ directly, we could try to evaluate the mean field $\langle f \rangle_T$ at a sufficiently high temperature and then track it down using the continuation method [32] as the temperature is lowered towards zero. In this connection, the analysis of the partition function is central; indeed, it is a central part in mean field theory because once it is calculated, all statistic information about the system can be deduced from it.

The partition function can be approximated by (1) rewriting the sum over a discrete space as an integral over a pair of continuous variables p and q, $p = \{p_i(I) \in \mathbb{R} \mid i \in \mathcal{S}, I \in \mathcal{L}\}$ and $q = \{q_i(I) \in \mathbb{R} \mid i \in \mathcal{S}, I \in \mathcal{L}\}$ having the same dimensionality as q, and then (2) evaluating its integrand at the saddle point (saddle point approximation) [27]. The mean field theory equations are derived as

$$q_i(I) = -\frac{1}{T} \frac{\partial E(p)}{\partial p_i(I)} \tag{30}$$

and

$$p_i(I) = \frac{e^{q_i(I)}}{\sum_J e^{q_i(J)}} \tag{31}$$

The derivation of mean field theory equations for combinatorial minimization can be found in many other places *e.g.* in [33], [31], [33], [5].

Note that in terms of the RL representation, the energy can be written as $E(f) = Const - G(f)$ and so q can be computed as

$$q_i(I) = \frac{1}{T}\left[r_i(I) + 2\sum_{i' \in \mathcal{S}}\sum_{I' \in \mathcal{L}} r_{i,i'}(I,I')\ p_{i'}(I')\right] \tag{32}$$

(*cf.* 19). Now we obtain the following fixed point iteration

$$p_i^{(t+1)}(I) \leftarrow \frac{e^{q_i^{(t)}(I)}}{\sum_J e^{q_i^{(t)}(J)}} \tag{33}$$

The p matrix is a labeling assignment in the feasibility space defined by (9). In iterative computation, the temperature T in q is decreased towards 0^+ as iteration goes. The unambiguity of p is achieved when $T \to 0^+$. In [27], the convergence is judged by a quantity defined as

$$S = \frac{1}{m}\sum_i\sum_I p_i^2(I) \tag{34}$$

also called the "saturation" of p. The iteration (33) is considered converged if $S > 0.99$. Analog network implementation schemes for mean field algorithms are investigated in [5].

2.5. Local vs. Global Solutions

Gradient-based optimization methods guarantees only local solutions when the cost function is non-convex. RL algorithms are local minimizers by nature. Given an RL iteration scheme and the data, two factors affect the solution: the initial assignment of labels and the compatibility function [30]. 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 [4], [24].

In some earlier RL works [30], [26], the dependence of solutions on initializations is regarded as a desirable property. Compatibility functions are defined *without* referring to the observations and therefore they should be considered as containing *a priori* information only. Constraints due to the observation are encoded in the initial assignment. Therefore, traditional RL solutions rely much on the initialization; convergence to a solution regardless of the initialization is considered undesirable. However, the posterior MRF energy already contains both sources of information, *i.e.* on the prior and the observation. Therefore the global energy minimum is the desirable solution [8].

2.6. Related Algorithms

Now we describe two other algorithms that will be compared experimentally: iterated conditional modes (ICM) [2] and simulated annealing (SA) [11], [8]. They are described for the following reasons: ICM is often used because of its fast convergence. On the other hand, Currently, the optimal solution is theoretically guaranteed only by a stochastic method like SA with a slow enough annealing schedule [8]. Thus, we may say that a bench mark for the comparison of all these approaches is the simulated annealing solution.

Iterated Conditional Modes Since it is difficult to maximize the joint posterior probability of an MRF, Besag [2] proposes the deterministic ICM algorithm which maximizes local conditional probabilities sequentially. The ICM algorithm uses the "greedy" strategy in the iterative local maximization. Given the data d and the other labels $f^{(k)}_{\mathcal{S}-\{i\}}$, the algorithm sequentially updates each $f_i^{(k)}$ into $f_i^{(k+1)}$ by maximizing $P(f_i \mid d, f_{\mathcal{S}-\{i\}})$, the conditional (posterior) probability, with respect to f_i.

Two assumptions are made in calculating $P(f_i \mid d, f_{\mathcal{S}-\{i\}})$: First, the observation components $d_1, \ldots, d_m$ are conditionally independent given f and each d_i has the same known conditional density function $p(d_i \mid f_i)$ dependent only on f_i. Thus

$$p(d \mid f) = \prod_i p(d_i \mid f_i) \tag{35}$$

The second assumption is that f depends on the labels in the local neighborhood, which is Marko-

initialize T and f;
repeat
    randomly sample f from $\mathcal{N}(f)$ under T;
    decrease T;
until ($T \to 0$);
return f;

Fig. 2. The simulated annealing algorithm.

vian. From the two assumptions and the Bayes theorem, it follows that

$$P(f_i \mid d, f_{\mathcal{S}-\{i\}}) \propto p(d_i \mid f_i)\ P(f_i \mid f_{\mathcal{N}_i}) \qquad (36)$$

Obviously, $P(f_i \mid d_i, f_{\mathcal{N}_i}^{(k)})$ is much easier to maximize than $P(f \mid d)$, which is the point of ICM.

Maximizing (36) is equivalent to minimizing the corresponding posterior potential using the following rule

$$f_i^{(k+1)} \leftarrow \arg\min_{f_i} V(f_i \mid d_i, f_{\mathcal{N}_i}^{(k)}) \qquad (37)$$

where

$$V(f_i \mid d_i, f_{\mathcal{N}_i}^{(k)}) = \sum_{i' \in \mathcal{N}_i} V(f_i \mid f_{i'}^{(k)}) + V(d_i \mid f_i) \qquad (38)$$

For discrete $\mathcal{L}$, $V(f_i \mid d_i, f_{\mathcal{N}_i})$ is evaluated with each $f_i \in \mathcal{L}$ and the label causing the lowest $V(f_i \mid d_i, f_{\mathcal{N}_i})$ value is chosen as the value for $f_i^{(k+1)}$. When applied to each i in turn, the above defines an updating cycle of ICM. The iteration continues until convergence. The convergence is guaranteed for the serial updating and is rapid [2].

The result obtained by ICM depends very much on the initial estimator $f^{(0)}$. Currently, it is not known how to set the initialization properly to obtain a good solution. A natural choice for $f^{(0)}$ is the maximum likelihood estimate which is the same as the data, $f^{(0)} = d$, when the noise is i.i.d. Gaussian.

Simulated Annealing SA, listed in Fig.2, is a stochastic algorithm for combinatorial optimization. It simulates the physical annealing procedure in which a physical substance is melted and then slowly cooled in search of a low energy configuration. SA applies a sampling algorithm, such as Metropolis algorithm [21], [11], or Gibbs sampler [8], successively at decreasing values of temperature T. Initially, T is set to a very high value and f is set to a random configuration. At a fixed T, the sampling is according to the Gibbs distribution $P_T(f) = \mathrm{e}^{-E(f)/T} / \sum_f \mathrm{e}^{-E(f)/T}$. After the Metropolis algorithm converges to the equilibrium at current T, T is decreased according to a carefully chosen schedule. This iterates until T is close to 0 when the system is "frozen" near the minimum of $E(f)$. At high T, large increases in E may be accepted; at lower T, only smaller increases may be accepted; and near the freezing T, no increases can be accepted. This enables the algorithm to escape from local minima into the global minimum.

Two convergence theorems are proven by [8]. The first concerns the convergence of the Metropolis algorithm. It says that if every configuration is visited infinitely often, the distribution of generated configurations is guaranteed to converge to the Boltzmann (*i.e.* Gibbs) distribution.

The second is about SA. It states that if the decreasing sequence of temperatures satisfy

$$\lim_{t\to\infty} T^{(t)} = 0 \qquad (39)$$

and

$$T^{(t)} \geq \frac{m \times \Delta}{\log(1+t)} \qquad (40)$$

where $\Delta = \max_f E(f) - \min_f E(f)$, then the system converges to the global minimum regardless of the initial configuration $f^{(0)}$. Note that the above conditions are sufficient but not necessary for the convergence.

Unfortunately, the schedule (40) is too slow to be of practical use. In practice, heuristic, faster schedules have to be used instead. Geman and Geman [8] adopt the following

$$T^{(t)} = \frac{C}{\log(1+t)} \qquad (41)$$

where the constant C is set $C = 3.0$ or $C = 0.4$ for their problem. Kirkpatrick *et al.* [11] choose

$$T^{(t)} = \kappa T^{(t-1)} \tag{42}$$

where κ typically takes a value between 0.8 and 0.99. The initial temperature is set high enough so that essentially all configuration changes are accepted. At each temperature, enough configurations are tried that either there are $10m$ (10 times the number of sites) accepted configuration transitions or the number of tries exceeds $100m$. The system is frozen and annealing stops if the desired number of acceptances is not achieved at three successive temperatures.

3. MAP-MRF Object Matching

Model-based object matching, or weighted graph matching in the abstract form, will be used as the test-bed for the experimental comparison of various algorithms. Let $\mathcal{S} = \{1, \ldots, m\}$ index the nodes in the first graph corresponding to the features in the scene and $\mathcal{L}^0 = \{1, \ldots, M\}$ index the nodes in the second graph corresponding to the features in the model. Introduce a virtual node called the NULL and indexed by 0 and augment $\mathcal{L}^0$ into $\mathcal{L} = \{0, 1, \ldots, M\}$. The NULL represents everything not included in $\mathcal{L}^0$, such as spurious features and features due to other model objects.

The scene and the model object are represented as graphs $\mathcal{G} = (\mathcal{S}, d)$ and $\mathcal{G}' = (\mathcal{L}, D)$, respectively, where $d = \{d_1, d_2\}$ and $D = \{D_1, D_2\}$ are feature descriptors composed of unary properties (d_1 and D_1) and binary relations (d_2 and D_2) [14]. The matching takes place from $\mathcal{G}$ to $\mathcal{G}'$ (or vice versa). The purpose is to find the correspondence between nodes of the two graphs so that the constraints on the node properties and the weights are best satisfied. An energy function measuring the cost of matching is formulated in the following based on Markov random fields.

3.1. Energy Function

In all cases, an $\mathcal{N}_i$ of the neighborhood system $\mathcal{N}$ on $\mathcal{S}$ can consist of all the other sites $i' \neq i$. This is a trivial case for MRF. In relational matching, it can consist of all the other sites which are related to i by any relations in d. When the scene is very large, $\mathcal{N}_i$ includes only those of the other sites which are within a spatial distance r from i

$$\mathcal{N}_i = \{i' \in \mathcal{S} \mid [\text{dist}(\text{feature}_{i'}, \text{feature}_i)]^2 \leq r,\ i' \neq i\} \tag{43}$$

The distance threshold r may be reasonably related to the size of the model object. The set of first order cliques is

$$\mathcal{C}_1 = \{\{i\} \mid i \in \mathcal{S}\} \tag{44}$$

The set of second order cliques is

$$\mathcal{C}_2 = \{\{i, i'\} \mid i' \in \mathcal{N}_i, i \in \mathcal{S}\} \tag{45}$$

Here, only cliques of up to order two are considered.

The definition of the prior potentials is a generalization of that penalizing line process variables [8], [19]. The single-site potential is defined as

$$V_1(f_i) = \begin{cases} v_{10} & \text{if } f_i = 0 \\ 0 & \text{otherwise} \end{cases} \tag{46}$$

where v_{10} is a constant. If f_i is the NULL label, it incurs a penalty of v_{10}; or the nil penalty is imposed otherwise. The pair-site potential is defined as

$$V_2(f_i, f_{i'}) = \begin{cases} v_{20} & \text{if } f_i = 0 \text{ or } \ f_{i'} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{47}$$

where v_{20} is a constant. If either f_i or $f_{i'}$ is the NULL , it incurs a penalty of v_{20} or nil penalty otherwise. The above clique potentials define the prior energy $U(f)$. The prior energy is then

$$U(f) = \sum_{i \in \mathcal{S}} V_1(f_i) + \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) \tag{48}$$

The conditional p.d.f., $p(d \mid f)$, of the observed data d, also called the *likelihood function*) when viewed as a function of f given d fixed, has the following characteristics:

1. It is conditioned on pure non-NULL matches $f_i \neq 0$,
2. It is independent of the neighborhood system $\mathcal{N}$, and
3. It depends on how the model object is observed in the scene which in turn depends on the underlying transformations and noise.

Assume (1) D and d are composed of types of features which are invariant under the considered class of transformations (The discovery and computation of visual invariants is an active area of research. See [23].); (2) they are related via the observation model

$$d_1(i) = D_1(f_i) + e_1(i) \quad (49)$$

$$d_2(i,i') = D_2(f_i, f_{i'}) + e_2(i,i') \quad (50)$$

where e_1 and e_2 are additive independent zero mean Gaussian noise. The assumptions of the independent and Gaussian noise may not be accurate but offers an approximation when an accurate observation model is not available.

Then the likelihood function is a Gibbs distribution with the energy

$$U(d \mid f) = \sum_{i \in \mathcal{S}, f_i \neq 0} V_1(d_1(i) \mid f_i) + \sum_{i \in \mathcal{S}, f_i \neq 0} \sum_{i' \in \mathcal{S}-\{i\}, f_{i'} \neq 0} V_2(d_2(i,i') \mid f_i, f_{i'}) \quad (51)$$

where the constraints, $f_i \neq 0$ and $f_{i'} \neq 0$, restrict the summations to take over the non-NULL matches. The likelihood potentials are

$$V_1(d_1(i) \mid f_i) = \begin{cases} \sum_{k=1}^{K_1} [d_1^{(k)}(i) - D_1^{(k)}(f_i)]^2 / \{2[\sigma_1^{(k)}]^2\} & \text{if } f_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (52)$$

and

$$V_2(d_2(i,i') \mid f_i, f_{i'}) = \begin{cases} \sum_{k=1}^{K_2} [d_2^{(k)}(i,i') - D_2^{(k)}(f_i, f_{i'})]^2 / \{2[\sigma_2^{(k)}]^2\} & \text{if } i' \neq i \& f_i \neq 0 \ \& \ f_{i'} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (53)$$

where $[\sigma_n^{(k)}]^2$ ($k = 1, \ldots, K_n$ and $n = 1, 2$) are the variances of the corresponding noise components. The vectors $D_1(f_i)$ and $D_2(f_i, f_{i'})$ are the "mean vectors", conditioned on f_i and $f_{i'}$, for the random vectors $d_1(i)$ and $d_2(i,i')$, respectively, conditioned on the configuration f.

Using $U(f \mid d) = U(f) + U(d \mid f)$, we obtain the posterior energy

$$U(f \mid d) = \sum_{i \in \mathcal{S}} V_1(f_i) + \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) + \sum_{i \in \mathcal{S}: f_i \neq 0} V_1(d_1(i) \mid f_i) + \sum_{i \in \mathcal{S}: f_i \neq 0} \sum_{i' \in \mathcal{S}-\{i\}: f_{i'} \neq 0} V_2(d_2(i,i') \mid f_i, f_{i'}) \quad (54)$$

This formulation makes use of contextual constraints not only on the prior configuration of labels but also on the observed data. It is a development based on previous works [13], [14]; there, an energy function for matching is proposed on a heuristic basis; here, (54) is formulated based on MRFs [15]. Another development from those works is made by [12] where the energy heuristically proposed in [13], [14] is derived from probabilistic viewpoint.

3.2. Energy-Gain Conversion

Substituting (46), (47), (52) and (53) into (54), we have

$$U(f \mid d) = v_{10} N_1 + v_{20} N_2 + \sum_{i \in \mathcal{S}: f_i \neq 0} V_1(d_1(i) \mid f_i) + \sum_{i \in \mathcal{S}: f_i \neq 0} \sum_{i' \in \mathcal{S}-\{i\}: f_{i'} \neq 0} V_2(d_2(i,i') \mid f_i, f_{i'}) \quad (55)$$

where $N_1 = \#\{f_i = 0 \mid i \in \mathcal{S}\}$ is the number of NULL labels in f, and $N_2 = \#\{f_i = 0 \text{ or } f_{i'} = 0 \mid i \in \mathcal{S}, i' \in \mathcal{N}_i\}$ is the number of label pairs at least one of which is NULL . The constants v_{10} and v_{20} control the number of sites assigned the NULL label. The smaller they are, the more sites will be assigned the NULL . The corresponding gain function is

$$G(f) = g_{10} N_1 + g_{20} N_2 + \sum_{i \in \mathcal{S}: f_i \neq 0} r_i(f_i) + \sum_{i \in \mathcal{S}: f_i \neq 0} \sum_{i' \in \mathcal{S}-\{i\}: f_{i'} \neq 0} r_{i,i'}(f_i, f_{i'}) \quad (56)$$

In the above, $g_{10} = Const_1 - v_{10}$, $g_{20} = Const_2 - v_{20}$, and $r_i(f_i)$ and $r_{i,i'}(f_i, f_{i'})$ are related to $V_1(d_1(i) \mid f_i)$ and $V_2(d_2(i,i') \mid f_i, f_{i'})$ by (16) and (17). These determines the compatibilities in RL. The gain is to be maximized.

3.3. *Annealing Labeling for MAP-MRF Matching*

Here we describe a special kind of graduated non-convexity (GNC) [3] algorithm, called annealing labeling [18], which may be used to improve the optimization of the posterior energy (54) formulated for the MAP matching.

The parameters g_{10} and g_{20} not only affect the NULL labels but also the local behavior of the maximization algorithm. When both are zero, there will be no NULL labels. The larger their values. the more sites will be assigned the NULL and the more serious the problem of local maxima becomes. In other words, we found that a labeling f, in which $f_i = 0$ for some i, is likely a local optimum and that the labeling f, in which $f_i = 0$ for all i, is a deep local optimum. This has motivated us to design a heuristic annealing procedure, similar to the GNC algorithm, to overcome the local optimum problem.

Introduce a temperature parameter into the gain

$$G(f \mid T) = \frac{1}{T}\{g_{10}N_1 + g_{20}N_2\} \\ + \sum_{i\in\mathcal{S}:f_i\neq 0} r_i(f_i) + \sum_{i\in\mathcal{S}:f_i\neq 0}\ \sum_{i'\in\mathcal{S}-\{i\}:f_{i'}\neq 0} r_{i,i'}(f_i, f_{i'}) \qquad (57)$$

with $G(f) = \lim_{T\to 1} G(f \mid T)$. T is initially set to a very high value $T^{(0)} \to \infty$ and gradually lowered towards $T^{(\infty)} = 1$. The maximum f^* obtained at $T^{(t-1)}$ is used as the initial value for the next new phase of maximization at $T^{(t)}$. In this way, the f^* are tracked from high T to low T. Because g_{10} and g_{20} are weighted by $\frac{1}{T}$, the optimum f^* obtained at high T is less affected by the local optimum problem; when it is tracked down, a better quality solution is usually obtained than one found by a non-annealing algorithm. The improvement of annealing labeling ICM, a procedure incorporating annealing labeling into ICM, over the simple ICM will be shown shortly.

Note that the T in the annealing RL acts on only the prior term whereas in SA, T acts on both the prior and likelihood terms. So the annealing RL is more like GNC [3] than SA [8].

4. Experiments

In the following experiments, we compare several minimization algorithms in terms of the solution quality measured by the maximized gain $G(f^*)$, the computational cost measured by the number of required iterations, and the need for heuristics for tuning the algorithm. The compared algorithms are:

1. ICM of Besag [2],
2. RL of Hummel and Zucker [10],
3. RL of Rosenfeld *et al.* [30],
4. MFA of Peterson and Soderberg [27] implemented with two schedules
5. SA of Kirkpatrick *et al.* [11],
6. Annealing ICM (Section 3.3).

Each value in the results shown below is obtained by averaging 200 runs of which each run finds a minimum f^* given a distinct pair of graphs.

The schedules for the annealing algorithms are as follows: For the SA, it is also $T^{(t+1)} \leftarrow 0.99T^{(t)}$ but with the initial temperature set to $10^9 \times M \times m$; annealing stops if the desired number of acceptances is not achieved after 100 successive temperatures [11]. These are purposely tuned for the best result with the least possible computational cost. The annealing ICM is quite insensitive to its schedule which is chosen as $T^{(t+1)} \leftarrow 0.9T^{(t)}$ with $T^{(0)} = 10$.

We implemented two annealing schedules for MFA. The first is the one originally given in [27] in which T is decreased according to $T^{(t+1)} \leftarrow 0.99T^{(t)}$; we set the initial temperature as $T^{(0)} = 10^9 \times M \times m$. The second is also in the form $T^{(t+1)} \leftarrow \kappa^{(t+1)}T^{(t)}$ but $\kappa^{(t+1)}$ is chosen on an *ad hoc* basis of trial and error to get the best result, which is $\kappa^{(t+1)} = 0.9$ if $S < 0.9$ & $\kappa^{(t)} \leq 0.9$, 0.95 if $S < 0.95$ & $\kappa^{(t)} \leq 0.95$, or 0.99 otherwise, where $S = \frac{1}{m}\sum_I p_i^2(I)$ is the saturation measure [27]. We refer to these two schedules as MFA-1 and MFA-2, respectively.

The value for μ parameter used in the GP operator of the Hummel-Zucker algorithm is chosen automatically in each iteration as the largest value subject to the inequalities in (25). Initially, $p^{(0)}$ is randomly set to a point in $\mathbb{P}$ (for continuous RL); $f_i^{(0)}(I)$ is randomly set to a number in $[0, 1, \ldots, M]$ (for discrete RL). The stopping criterion for the continuous RL algorithms (Hummel-Zucker, Rosenfeld *et al.* , and Peterson and Soder-

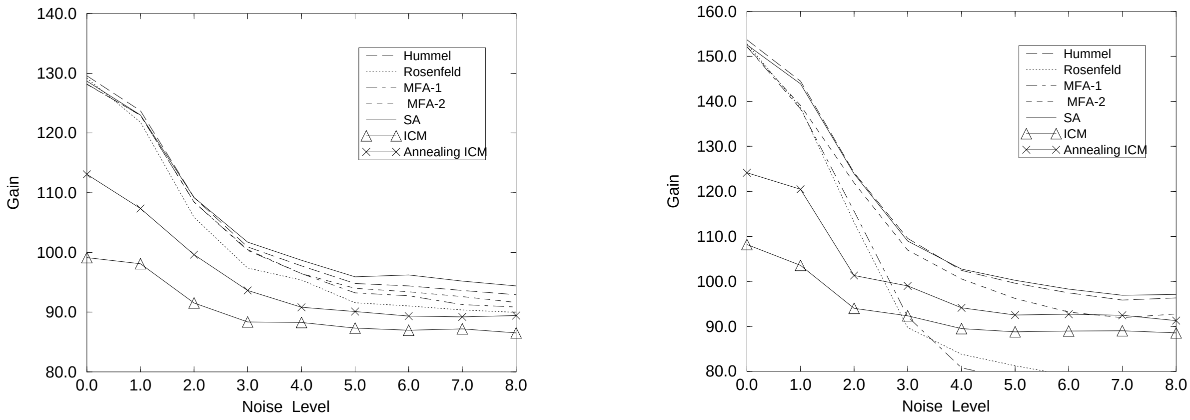

Fig. 3. The maximized gain $G(f^*)$ (after divided by m^2), for $m = 10$ (top) and $m = 20$ (bottom). The higher, the better.

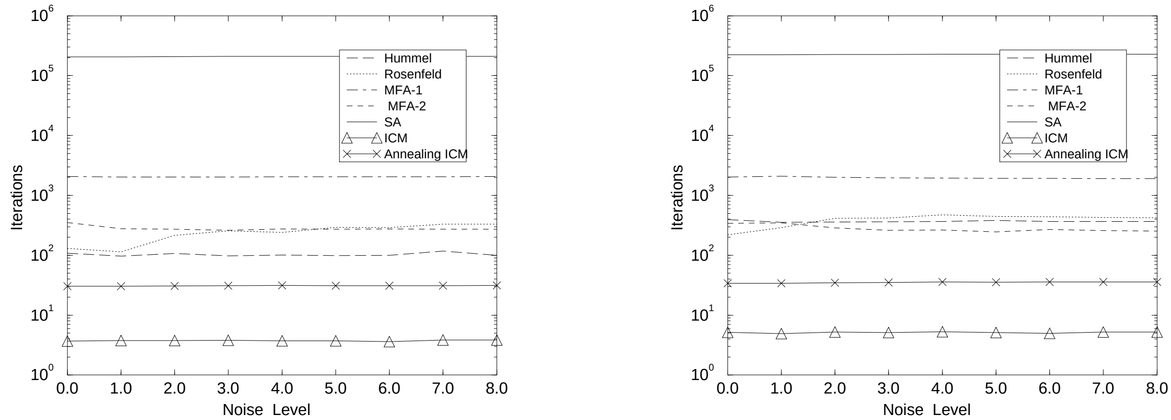

Fig. 4. The number of iterations, for $m = 10$ (top) and $m = 20$ (bottom). The lower, the better.

berg) is when the saturation S defined in (34) is smaller than, say, 0.001. The stopping criterion for ICM is that no more state changes occur.

The test bed is the MRF matching of weighted graphs. Here, a node corresponds to a point in $X-Y$ plane. Two random graphs, containing $m = M$ nodes, respectively, are randomly generated as follows. First, a number of $\lfloor \frac{2}{3}m \rfloor$ nodes, where $\lfloor \cdot \rfloor$ is the "floor" operation, are generated using random numbers uniformly distributed within a box of size 100×100. They are for the first graph. Their counterparts in the other graph are generated by adding Gaussian noise of $N(0, \sigma^2)$ to the x and y coordinates of each of the nodes in the first graph. This gives $\lfloor \frac{2}{3}m \rfloor$ deviated locations of the nodes for the other graph. Then, the rest of the $\lceil \frac{1}{3}m \rceil$ nodes in each of the two graphs are generated using random numbers uniformly distributed in the box. This simulates outlier nodes.

The two sets of nodes are indexed by $\mathcal{S}$ and $\mathcal{L}$, respectively. The only constraint used for matching is the distance between nodes. The distances constitute the relations in d and D. The two graphs are represented as $\mathcal{G} = (\mathcal{S}, d)$ and

$\mathcal{G}' = (\mathcal{L}, D)$. The matching takes place from $\mathcal{G}$ to $\mathcal{G}'$ (or vice versa). When the noise variance is 0, there must be $\lfloor \frac{2}{3} m \rfloor$ exact matches though other matches due to coincidence are not impossible. When noise is present, the matching becomes inexact.

The posterior energy formulated in Section 3 is used to measure the cost of a matching $f = \{f_1, \ldots, f_m\}$. The energy is converted to the gain to be maximized (*cf.* Sections 2.3 and 3.2). To reduce the demand on storage, we used only one byte (8 bits) to represent the compatibility coefficients $r_{i,i'}(I, I')$ in the RL formula. Positive values are truncated to an integer between 0 and 255 while negative values are truncated to zeros. In this case, we set $Const_2 = 255$ in calculating the pair-wise compatibilities (17). Although the precision of the compatibility coefficients is low, good results are still obtained; this demonstrates the error-tolerant aspect of the RL minimization approach.

Fig.3 illustrates the solution quality in terms of the maximized gain as a function of the noise level σ. The Hummel-Zucker algorithm performs well all the time. The SA presents its ability when noise is large. MFA-2 is just behind the above two algorithms. MFA-1 and the original Rosenfeld algorithm perform satisfactorily when $m = M = 10$; however, when $m = M = 20$ and higher, the results deteriorate quickly as σ^2 increases and drop to even lower than those of the two ICM algorithms. Looking at the MFT updating rule (33), we find that (33) does not necessarily lead to a decrease in $E(f)$ or increase in $G(f)$; and we observed that at low T, and we observed that the original MFA may collapse to a low quality f^* on convergence. The schedule used in MFA-2 seems overcome this problem. Lastly, we see that adding the annealing procedure to ICM significantly improves the quality of the ICM solution.

Fig.4 demonstrates the computational cost in terms the number of iterations. ICM converges very fast, after just a few iterations. In contrast, the number for SA with the specified heuristic schedule is several order higher than all the deterministic algorithms. The other algorithms are in between.

In [28], the evaluation of RL algorithms is in terms of the number of iterations and the number of errors. The goodness of interpretation used in [28] not only relates to algorithms themselves but also to how the problem is formulated (*e.g.* the definition of compatibilities), mixing the two issues, problem formulation and computational algorithm. In this paper, we measure the solution quality by the quantitative gain $G(f^*)$ and regard RL as just a mechanism of minimization rather than one of interpretation. The concern here is how well an algorithm can minimize an energy regardless of the problem domain. Price concludes is that the algorithms of Faugeras and Price [7] and Hummel and Zucker [10], which use the gradient projection techniques, are the best — a result consistent with ours; that by Peleg [25] converges too fast to yield good result; and that of Rosenfeld *et al.* performs only adequately. There, the numerical performance measures are obtained using a few runs; in contrast, our results are obtained after a much larger number of runs and should be statistically more reliable.

5. Conclusions

In this paper, relations between relaxation labeling and the minimization of MRF energy with discrete labels are established. Various algorithms are compared in terms of maximized gain values and required numbers of iteration. The comparison leads to the following conclusions: Simulated annealing gives good quality solutions when the problem size but is time consuming. The Hummel-Zucker algorithm yields good quality solutions, comparable to SA, and requires much less computational effort than SA, well balancing between the quality and the cost. Further, it avoids cumbersome needs for the heuristic tuning of annealing schedules in annealing algorithms. MTF annealing would also be a good choice when the rapid deterioration of solutions for "harder" problems is remedied for example by using an improved annealing schedule. ICM converges very fast but does not produce good quality solutions.

Besides the solution quality and the computational cost, an important factor that must be taken into account is the need for, and difficulties in, the tuning of the annealing schedule. The scheduling is critical to the success of SA and is an area of study [1] and it also affect very much the mean field algorithm. How to choose an optimal schedule depends not only on the type, but also on the size, of the problem. Finding a good heuristic schedule can be tedious. A slower schedule does not necessarily yield better results.

Current continuous RL algorithms such as those mentioned in this paper are parallelizable but not distributed. The normalization is needed so that the assignments $p_i(I)$ over all I are summed to 1. In a recent paper [17], RL is solved as a constrained optimization problem by using Lagrange multipliers together with a technique in the graded Hopfield neural network. This leads to a continuous RL algorithm which is fully parallel and distributed and may be implemented with analog circuits.

References

1. E. H. L. Aarts. *Simulated annealing and Boltzmann machines : a stochastic approach to combinatorial optimization and neural computing.* Wiley, 1989.
2. J. Besag. "On the statistical analysis of dirty pictures" (with discussions). *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.
3. A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, Cambridge, MA, 1987.
4. 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.
5. I. M. Elfadel. *From Random Fields to Networks.* PhD thesis, Department of Mechanical Engineering, 1993.
6. 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*, 3:412–423, April 1981.
7. O. D. Faugeras and K. Price. "Semantic description of aerial images using stochastic labeling". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3:638–642, November 1981.
8. S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.
9. R. M. Haralick. "Decision making in context". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 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*, 5(3):267–286, 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. " Towards 3D vision from range images: An optimization framework and parallel networks". *CVGIP: Image Understanding*, 55(3):231–260, May 1992.
14. S. Z. Li. "Matching: invariant to translations, rotations and scale changes". *Pattern Recognition*, 25(6):583–594, June 1992.
15. 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*, pages 866–869, Seattle, Washington, June 1994.
16. S. Z. Li. *Markov Random Field Modeling in Computer Vision.* Springer-Verlag, New York, 1995.
17. S. Z. Li. "Relaxation labeling using Lagrange multipliers and Hopfield network". In *Proceedings of IEEE International Conference on Image Processing*, Washington, D.C., 23-26 October 1995.
18. S. Z. Li, H. Wang, and M. Petrou. "Relaxation labeling of Markov random fields". In *Proceedings of International Conference Pattern Recognition*, volume 1, pages 488–492, Jerusalem, Israel, October 1994.
19. J. L. Marroquin. *Probabilistic Solution of Inverse Problems.* PhD thesis, MIT AI Lab, 1985.
20. B. M. McMillin and L. M. Ni. "A reliable parallel algorithm for relaxation labeling". In P. M. Dew, R. A. Earnshaw, and T. R. Heywood, editors, *Parallel Processing for Computer Vision and Display*, pages 190–209. Addison-Wesley, 1989.
21. N. Metropolis and *etal.* "Equations of state calculations by fast computational machine". *Journal of Chemical Physics*, 21:1087–1091, 1953.
22. J. Mohammed, R. Hummel, and S. Zucker. "A feasible direction operator for relaxation method". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(3):330–332, May 1983.
23. J. L. Mundy and A. Zisserman, editors. *Geometric Invariants in Computer Vision.* MIT Press, Cambridge, MA, 1992.
24. 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.
25. S. Peleg. "A new probability relaxation scheme". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:362–369, 1980.
26. S. Peleg and R. A. Rosenfeld. "Determining compatibility coefficients for curve enhancement relaxation processes". *IEEE Transactions on Systems, Man and Cybernetics*, 8:548–554, July 1978.
27. C. Peterson and B. Soderberg. "A new method for mapping optimization problems onto neural networks". *International Journal of Neural Systems*, 1(1):3–22, 1989.
28. K. E. Price. "Relaxation matching techniques – A comparison". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(5):617–623, September 1985.

29. S. A. Rosen. "The gradient projection method for nonlinear programming — part i: Linear constraints". *Journal of the Society for Industrial and Applied Mathematics*, 8(1):181–217, 1960.
30. A. Rosenfeld, R. Hummel, and S. Zucker. "Scene labeling by relaxation operations". *IEEE Transactions on Systems, Man and Cybernetics*, 6:420–433, June 1976.
31. P. D. Simic. "Statisrical mechanics as the underlying theory of 'elastic' and 'neural' optimization". *Network*, 1:89–103, 1990.
32. E. Wasserstrom. Numerical solutions by the continuition method". *SIAM Review*, 15:89–119, 1973.
33. A. L. Yullie. "Generalized deformable models, statistical physics and matching problems". *Neural Computation*, 2:1–24, 1990.

Biography

Stan Z. Li received the B.Eng degree from Hunan University, China, in 1982, M.Eng degree from the National University of Defense Technology, China, in 1985 and Ph.D degree from the University of Surrey, UK, in 1991. All degrees are in EEE. His research interests include computer vision, pattern recognition, image processing and optimization methods. He is the author of the book *Markov Random Field Modeling in Computer Vision* (Springer-Verlag, 1995).

Han Wang received his B.Eng in computers and its applications in 1982 from Northeast Heavy Machinery Institute, China and his Ph.D in Parallel Image Processing in 1990 from Leeds University, UK. He has been a visiting scientist (1988) in Carnegie Mellon University for developing a parallel image processing language for the Warp machine and for Transputers. From 1989 to 1992, he has been a research officer in the Robotics Research Group and the Computing Laboratory, Oxford University. The research involved in real time implementation of 3D image processing algorithms for mobile robot navigation. Dr. Wang is currently lecturing in the Nanyang Technological University and his research interests are in object based real-time tracking and determination of hand pose from image sequences that can be used for control of a 5 degree-of-freedom robot head.

Kap Luk Chan received the BEng degree in Electronic Engineering in 1987 and the PhD degree in Robot Vision in 1991, all from the University of London, England. From 1990 to 1992, he worked as an Information Engineer in Axiom System Design, London, U.K. He joined the Nanyang Technological University, Singapore as a lecturer in the School of Electrical and Electronic Engineering in 1992. His research interests are in image analysis and robotic vision.

Maria Petrou received her B.Sc. in Physics in 1975 from the University of Thessaloniki, Greece, and her Ph.D. in Astronomy in 1981 from the University of Cambridge, United Kingdom. She has been working on Computer Vision since 1986, and has published more than 120 papers, more than 40 of which are in refereed journals, on Astronomy, Low Level Vision, Feature Extraction, Texture Analysis, Markov Random Fields, Probabilistic Relaxation, Remote Sensing, Industrial Inspection etc. She is a Senior lecturer at the Department of Electronic & Electrical Engineering of Surrey University, United Kingdom, and a member of the British Machine Vision Association, the Society of Optical Engineering and of IEEE. She is Associate Editor of IEEE Transactions on Image Processing, and the Editor of the Newsletter of the International Association of Pattern Recognition.