Relational Matching with Stochastic Optimisation

Andrew D.J. Cross and Edwin R. Hancock
Department of Computer Science, University of York
York, Y01 5DD, UK

Abstract

This paper describes a novel framework for performing relational graph matching by stochastic optimisation. The starting point for this study is a configurational probability measure which gauges the consistency of relational matches using a compound exponential function of Hamming distance. In order to overcome some of the well documented shortcomings of deterministic updating, we develop two contrasting stochastic optimisation strategies. The first of these exploits the apparatus of statistical physics to compute the Boltzmann distribution that models the configurational probability measure so that relational matching may be performed by simulated annealing. The second approach is a genetic hill climbing algorithm which is motivated by the way in which our configurational probability measure models matching errors. Because the genetic optimisation commences with a pool of random matches, it obviates the need for accurate initialisation. Moreover, it can recover consistent matches exploiting only structural information from the graphs under match.

1 Introduction

Discrete relaxation [4, 10] is a constraint satisfaction process that is concerned with assigning symbolic interpretations to arrangements of objects abstracted in terms of relational graphs. The process is invariably realised by parallel iterative computation aimed at optimising an objective function. This objective function grades the quality of interpretation by measuring the degree to which symbolic constraints operating in the different relational graphs are violated. Usually, however, the objective function is goal directed [8] and is only weakly grounded in information theory. For instance, in the classical pattern recognition literature Shapiro and Haralick [8] have proposed a relational distance metric which counts constraint violations at the clique level. Sanfeliu and Fu [9], on the other hand, have a finer definition of relational distance which counts the number of symbolic operations required to transform graphs into one-another. Wong and You [11] have taken some steps towards extending these ideas into the information theoretic domain by using probabilistic and entropic ideas to model the structural disturbance of noise corrupted graphs.

We have recently enhanced the available methodology by developing a Bayesian framework for relational matching by discrete relaxation [4, 5, 10]. We commence by specifying the matching process in terms of probability distributions which model relational constraint violations using a memoryless error-process. Moreover, this simple yet effective model naturally leads to a Bayesian objective function which gauges the quality of match by a series of exponential functions of the Hamming distances between matched relations and their idealised counterparts. In fact, in the case of binary patterns and when the probability of label errors is approximately equal to $\frac{1}{2}$, our discrete relaxation process becomes identical in function to the Hopfield memory [5]. Hitherto, the optimisation of this Bayesian objective function has been confined to the use of basic gradient ascent methods [10]. Although demonstrably effective, the technique is clearly prone to problems associated with convergence to a local optimum. A more satisfactory approach would be to locate the globally optimal graph match by some form of stochastic optimisation [1, 3, 2].

We study two contrasting stochastic optimisation methods. The first of these involves exploiting the apparatus of statistical physics [12] to compute effective Gibbs potentials so that our discrete relaxation process can be mapped onto an equivalent Boltzmann machine [1]. With the Gibbs potential to hand optimal updates may be located by simulated annealing using the well known Metropolis algorithm [1, 3]. Viewed from the perspective of both classical relational graph-matching [8, 9, 11] and statistical physics approaches [12] this aspect of our work can be regarded as providing a more objective framework for determining effective energy functions. The second stochastic optimisation process is based on genetic hill climbing [2]. This optimisation process draws more directly on the relational error model that underpins our matching framework. The technique commences with a population of randomly initialised matches. Here the Bayesian objective function plays the role of fitness measure, while random matching mutations occur at the prevailing value of the label-error probability. Because the matching process commences from a random population, initialisation ceases to be an important issue. Operating in conjunction with genetic hill-climbing, the Bayesian objective function is capable of imposing global consistency using only structural information from the graphs under match. Moreover, by maintaining a population of weighted matched graphs, we naturally bridge the methodological gap between discrete relaxation [4] and continuous labelling techniques such as probabilistic relaxation or mean field theory [7].

0-8186-7190-4/95 $4.00

The outline of this paper is as follows. In Section 2 we review our Bayesian framework for discrete relaxation. Section 3 describes our framework for computing equilibrium Gibbs potentials. Section 4 describes the different optimisation techniques that have been used to locate optimal matches. Comparative experimental evaluation of these optimisation techniques is presented in Section 5. Finally, Section 6 presents some conclusions.

2 Discrete Relaxation

Discrete relaxation is concerned with matching relational graphs represented in terms of configurations of symbolic labels. We represent such a graph by $G = (V, E)$, where V is the symbolic label-set assigned to the set of nodes and E is the set of edges between the nodes. Formally, we represent the matching of the nodes in the data graph $G_1 = (V_1, E_1)$ against those in the model graph $G_2 = (V_2, E_2)$ by the function $f : V_1 \rightarrow V_2$. In other words, the current state of match is denoted by the set of Cartesian-pairs constituting the function f.

In order to describe local interactions between the nodes at a manageable level, we will represent the graphs in terms of their clique structure. The clique associated with the node indexed j consists of those nodes that are connected by an edge of the graph,i.e. $C_j = \{i \in V_1 | (i, j) \in E_1\}$. The labelling or mapping of this clique onto the nodes of the graph G_2 is denoted by $\Gamma_j = \{f(i) \in V_2, \forall i \in C_j\}$. Suppose that we have access to a set of patterns that represent feasible relational mappings between the cliques of graph G_1 and those of graph G_2. Typically, these relational mappings would be configurations of consistent clique labellings which we want to recover from an initial inconsistent state of the matched graph G_1. Assume that there are Z_j relational mappings for the clique C_j which we denote by $\Lambda^\mu = \{\lambda_i^\mu \in V_2, \forall i \in C_j\}$ where $\mu \in \{1, 2...Z_j\}$ is a pattern index. According to this notation $\lambda_i^\mu \in V_2$ is the match onto graph G_2 assigned to the node $i \in V_1$ of graph G_1 by the μ^{th} relational mapping. The complete set of legal relational mappings for the clique C_j are stored in a *dictionary* which we denote by $\Theta_j = \{\Lambda^\mu | \mu = 1, Z_j\}$.

The discrete relaxation procedure is based on maximising the joint probability of the matched label configuration, i.e. $P(\Gamma_j)$. It is therefore necessary to find a way of enumerating $P(\Gamma_j)$ when the label configuration is highly inconsistent, i.e. when there are no dictionary items for which the Hamming distance is zero. The approach is to adopt a Bayesian viewpoint in which it is assumed that only consistent labellings in the dictionary are legal and have uniform non-zero *a priori* probabilities of occurrence, i.e. $P(\Lambda^\mu) = Z_j^{-1}$. Other configurations do not occur *a priori* but are the corrupted realisations of the dictionary items. This idea is realised by applying the axiomatic property of joint probability to expand $P(\Gamma_j)$ over the space of consistent configurations

$$P(\Gamma_j) = \sum_{\mu=1}^{Z_j} P(\Gamma_j | \Lambda^\mu) P(\Lambda^\mu) \tag{1}$$

Further development of a useful objective function for discrete relaxation requires a model of the label corruption process, that is of the conditional probabilities of the potentially inconsistent configurations given each of the Z_j feasible relational mappings $P(\Gamma_j | \Lambda^\mu)$. We adopt a very simple viewpoint; matching errors are assumed to be memoryless and to occur with uniform probability p.

The first consequence of the assumed absence of memory is that the errors are independent. As a result we can we can factorize the conditional probabilities over the individual nodes in the graph, i.e.

$$P(\Gamma_j | \Lambda^\mu) = \prod_{i \in C_j} P(f(i) | \lambda_i^\mu) \tag{2}$$

Our next step is to propose a model for the label corruption mechanism at each node in the graph. Again, taking recourse to the memoryless assumption, the probability of label-errors on individual objects is independent of the class of label. This leads us to the following assignment of probability

$$P(f(i) | \lambda_i^\mu) = \begin{cases} \text{1-p} & \text{if } f(i) = \lambda_i^\mu \\ \text{p} & \text{otherwise} \end{cases} \tag{3}$$

As a result of this distribution rule, the conditional matching probabilities depend on the Hamming distance H_μ between the matched configuration Γ_j and the individual dictionary items Λ^μ, i.e.

$$P(\Gamma_j | \Lambda^\mu) = (1 - p)^{|C_j| - H_\mu} p^{H_\mu} \tag{4}$$

where the Hamming distance H_μ is defined to be $H_\mu = \sum_{i \in C_j} (1 - s_{i, \lambda_i^\mu})$. The model components given in equations (2), (3) and (4) naturally lead to the following expression for $P(\Gamma_j)$ in terms of the set of Hamming distances to the consistent labellings residing in the dictionary

$$P(\Gamma_j) = \frac{b}{Z_j} \sum_{\mu=1}^{Z_j} \exp[-\beta H_\mu] \tag{5}$$

where $b = (1 - p)^{|C_j|}$ and $\beta = \ln \frac{1-p}{p}$. According to our picture of discrete relaxation, Hamming distance is the basic measure of consistency. Systematic softening of the constraints residing in the dictionary is controlled by the parameter p. It is tempting to draw analogies between the exponentials appearing in equation (5) and the Boltzmann distribution. The quantity β clearly plays the role of inverse temperature while H_μ is related to the Gibbsian potential. However, as we shall see in Section 3, the specification of the equivalent Boltzmann distribution is rather more subtle than this initial superficial analysis reveals.

The configurational probability $P(\Gamma_j)$ is the basic ingredient of our discrete relaxation scheme. It represents the probability of a particular matching configuration evaluated over the state-space of feasible possibilities (i.e. the dictionary). Since our model of the state-space demands that configurations outside the dictionary occur with zero probability, $P(\Gamma_j)$ plays the role of local partition function for the clique C_j. This is an important observation. It means that we can apply the apparatus of statistical physics [7, 12] to $P(\Gamma_j)$ to understand the collective behaviour of our discrete relaxation process.

In the next section we will describe a methodology for deriving an equivalent Boltzmann distribution that can be used to represent the configurational probability $P(\Gamma_j)$ appearing in equation (5). With this distribution to hand we can locate matches by simulated annealing, avoiding potential problems associated with convergence to a local minimum.

3 Effective Gibbs Potential

In order to realise relational graph matching by simulated annealing [1, 3], we require a Boltzmann distribution that models the configurational probability measure described in the previous Section. To this end we draw on the techniques of statistical physics [7, 12] to compute an equivalent Gibbs potential. Since we commenced our development of the matching probability in equation (5) by specifying the nature of the constraint corruption process, this statistical physics treatment provides an objective methodology for computing configurational potentials for inexact graph matching; this is to contrasted with the goal directed approach described in [6]. Moreover, since the underlying constraint corruption process is itself relatively simple, the construction of a model of the state-space of discrete relaxation is a feasible proposition. Consequently, we can explicitly model how the mean-value of the equilibrium potential evolves as the parameter p is reduced to impose consistency in the annealing schedule.

We commence our investigation by noting that if we represent the discrete relaxation process by an equivalent Boltzmann distribution with an identical partition function, then according to statistical physics the equilibrium Gibbs potential is given by [7]

$$U(\Gamma_j) = -\frac{1}{P(\Gamma_j)}\frac{\partial P(\Gamma_j)}{\partial \beta} \tag{6}$$

Upon substituting for $P(\Gamma_j)$ from equation (5),

$$U(\Gamma_j) = \frac{\sum_{\mu=1}^{Z_j} H_\mu \exp[-\beta H_\mu]}{\sum_{\mu=1}^{Z_j} \exp[-\beta H_\mu]} \tag{7}$$

In other words, the effective potential for a particular label configuration is just a weighted sum of Hamming distances over the different dictionary items. In fact, the exponential weights naturally suppress contributions from dictionary items of large Hamming distance. In other words, the potential is concentrated at intermediate values of Hamming distance associated to noise corruption. Large Hamming distances (of the order $|C_j|$) are associated with comparisons of different valid dictionary items and do not contribute significantly to the update process. This feature limits the disruptive effects of inter-pattern competition when the dictionary is large, i.e. there are many feasible relational mappings between the cliques of the graphs under match. In the context of binary memories, we have already shown how the exponential character of our discrete relaxation process leads to both vastly improved storage capacity and to enhanced pattern reconstruction capabilities when compared to the Hopfield network [5].

Before proceeding to issues of optimisation, it is interesting to consider how the Gibbs potential evolves as the relational matching process iterates to convergence. By drawing on a very simple model of the pattern-space of the relational matching process, it is possible to show that mean value of the Gibbs potential at the epoch when the coupling constant takes on the value β is

$$< U >_\beta = \frac{|C_j| \exp[-2\beta]}{1 + \exp[-2\beta]} \tag{8}$$

As a result in the low temperature limit as $\beta \rightarrow \infty$, the mean effective potential approaches zero. In other words, our Hamming distance potentials ensure that the ground state of our Boltzmann distribution is one of zero energy.

4 Optimisation

4.1 Deterministic Hill Climbing

Hitherto [5, 10], we have located relational matches using a simple hill-climbing approach to maximise the sum of clique configurational probabilities i.e.

$$P_G = \frac{1}{|V_1|}\sum_{j \in V_1} P(\Gamma_j) \tag{9}$$

This deterministic update process proceeds in a parallel iterative fashion commencing from an initial estimate of the optimal match; this initial match is established on the basis of the similarity of unary measurements between nodes of the two graphs. In order to impose consistency on the recovered match, the parameter p is incrementally reduced as the number of iterations increases. This is analogous to temperature annealing. As $p \rightarrow 0$ or equivalently $\beta \rightarrow \infty$, the exponentials appearing in equation (5) approach their Dirac delta-function limits. Viewed from the perspective of the state-space of relational mappings, the gradual reduction of p has the effect of incrementally hardening the available constraints residing in the dictionary.

4.2 Simulated Annealing

Having established the equivalent Boltzmann distribution for our discrete relaxation process, $Q(\Gamma_j)$, we may justifiably perform simulated annealing on the sum of the configurational clique potentials, i.e.

$$U_G = \sum_{j \in V_1} U(\Gamma_j) \tag{10}$$

Using the Metropolis algorithm [1, 3], we randomly select both nodes and potential updates. The update is accepted if it leads to a reduction in the global configurational potential U_G. If this is not the case, then the random update is accepted with probability

$$P_r = \frac{\exp\left[-\beta \sum_{i \in C_j} U(\Gamma_i)\right]}{\sum_{\Gamma_i} \exp\left[-\beta \sum_{i \in C_j} U(\Gamma_i)\right]} \tag{11}$$

The initial matches are established in an identical manner to those used in the deterministic update process.

4.3 Genetic Hill Climbing

Genetic search [2] is in many ways a more natural means of locating the optima of our configurational probability measure. In essence the approach relies on generating a population of random global matching configurations. These undergo crossover, mutation and selection to locate the match that optimises a fitness measure. The mutation phase of the update process is in many ways analogous to the Metropolis algorithm. However, since the algorithm commences from a set of random matches, accurate initialisation ceases to be of limiting importance. By incorporating a deterministic hill-climbing stage into the algorithm we accelerate convergence. This acceleration is applied to the fitness measure once mutations have occurred and is used to discourage gross inconsistencies entering the matching process.

The main stages of the genetic hill-climbing algorithm are as follows. Firstly, pairs of global matching configurations are randomly selected from the current population and random matches are interchanged with uniform probability $\frac{1}{2}$. A further randomisation stage is then applied to the individual matches to introduce new information into the population of global matches through a process of mutation. This is effected by updating the matches at random with probability p, i.e. the prevailing value of the label-error probability; this is the only parameter of our method, which fulfills the role of control variable. The hill-climbing and selection stages of the algorithm are aimed at optimising the global configurational probability measure P_G. Suppose that $P_G^{(i)}$ denotes the global configurational probability for the i^{th} member of the pool (population) of graphs. The hill-climbing stage involves reconfiguring each of the randomised matches using the deterministic update process described in Section 4.1 to maintain consistency. This not only accelerates convergence, it also diminishes the requirements for a large population of graphs. These optimised configurations are randomly admitted to the pool of graphs $\mathcal{P}$ with probability

$$P_s = \frac{P_G^{(i)}}{\sum_{i \in \mathcal{P}} P_G^{(i)}} \tag{12}$$

This process is iterated while incrementally reducing the value of the label-error probability p which controls the mutation rate. The final optimal match is located by selecting the graph for which $P_G^{(i)}$ is maximum. The idea of maintaining a population of alternative weighted matching configurations effectively bridges the conceptual gap between classical discrete relaxation methods [4] and continuous labelling algorithms such as probabilistic relaxation or mean-field annealing [12].

5 Experiments

Our experimental study consists of two parts. The first of these is aimed at demonstrating the relative merits of the different optimisation strategies on simulated graphs. The aim of this part of the study is to establish the operational limits of the deterministic update process when compared with the two stochastic update procedures. Experiments on real data aim to demonstrate two advantages of the genetic update process. The first of these is the fact that because genetic search commences from a pool of random matching configurations, accurate initialisation is not an issue of central importance. The second aim is to demonstrate that the genetic method outperforms both simulated annealing and deterministic hill-climbing in terms of its ability to converge rapidly on a global optimum.

5.1 Synthetic Data

We commence our experimental study by comparing the stochastic optimisation of the global effective potential U_G with the deterministic optimisation of the configurational probability P_G under conditions of controlled initialisation error. To compare the two optimisation processes, we have investigated the matching of randomly generated graphs containing controlled levels of corruption. Figure 1 shows the final fractional accuracy as a function of initial random corruption. The upper curve is the result of stochastic optimisation of the effective Gibbs potential, while the lower curve shows result obtained by deterministic optimisation of the configurational probability. Although both optimisation algorithms perform well at low corruption levels, once the fraction of initialisation error exceeds 40%, the deterministic method becomes trapped in a local optimum and fails to find a globally consistent match.

Even when the initialisation error is 100%, the simulated annealing method is capable of locating a globally consistent match. Because the genetic search algorithm commences from a pool of random matches, it is not possible to display the results in a way that is directly comparable with the curves in Figure 1. However, when presented with identical data, the genetic search technique was capable of recovering a globally consistent match in only 5-10 iterations.

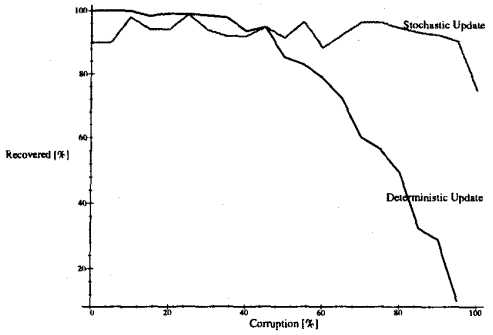

Figure 1: Comparison of stochastic and deterministic update

5.2 Real-world imagery

To provide some indication as to the relative performance of the three optimisation strategies, we have conducted some experiments on the matching of aerial images against digital map data. In this application the image data is rotated, scaled and distorted with respect to the digital map. At the level of image features, we are interested in matching road networks detected in the image data against their corresponding map representation. The image data under study was obtained using an infrared line scan sensor. In addition to the transformational differences due to changes of viewpoint and scale, this sensor introduces sources of variability due to scanning distortions at the image periphery. An example image is shown in the lefthand of Figure 2 while the righthand image is the digital map data available for matching. The righthand image in Figures 5 and 6 represents the road network segmented from the image data. Our matching of the two scenes is based on finding correspondences between the T-junctions and line-endings which delineate the road network. According to our graph-based abstraction of the matching process the nodes represent line-endings or T-junctions while the arcs signify the existence of a connecting line-structure.

To illustrate the relative convergence performance for the three optimisation strategies, Figure 3 shows a plot of P_G as a function of iteration number. In the case of the genetic algorithm the value plotted is the maximum configurational probability evaluated over the population of matches. Although deterministic hill-climbing has rapid and uniform convergence, the final value of P_G is suboptimal when compared to the alternatives. Simulated annealing, on the other hand,

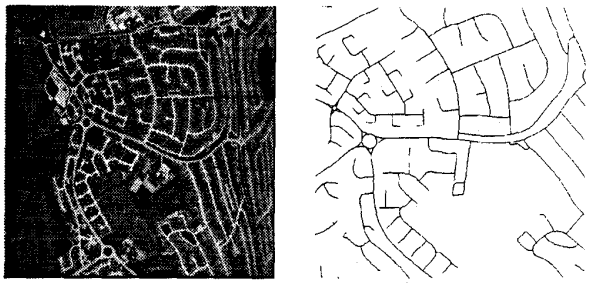

Figure 2: Aerial infra-red image (left) and digital map (right)

converges to a larger value of P_G, but is slow and non-uniform. Genetic search converges rapidly and uniformly to the maximum value of P_G.

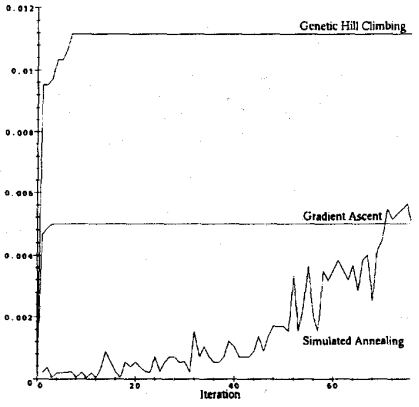

Figure 3: Iterative convergence

It is instructive to illustrate some of the iterative properties of our genetic search method. Figure 4 shows population histograms of P_G at different epochs of genetic search. This plot shows two important features. Firstly, the mode of the histogram moves rapidly from close to zero to the maximum value of P_G located by deterministic hill climbing (see Figure 4). Secondly, as time proceeds the high fitness tail of the histogram extends to the global maximum of configurational probability. It is this feature that ensures rapid convergence to a global optimum. By contrast, the deterministic update process remains trapped at the mode of the P_G histogram.

Figures 5 and 6 illustrate the matching results with the deterministic version of our discrete relaxation algorithm. Correct matches are shown in Figure 5 while matching errors are contained in Figure 6. Initially, of the 158 T-junctions and line-endings in the map 35 match correctly. After application of the discrete relaxation method, 72 match correctly; matches has therefore increased from 0.22 to 0.46. The fraction of nodes for which a feasible match exists is 0.54. In the case of simulated annealing the fraction of nodes correctly matched is 0.48, while in the case of genetic update it is 0.51. However,

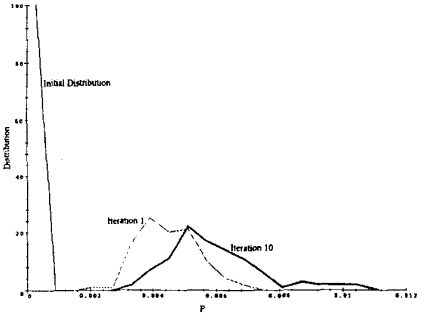

Figure 4: Population histograms for genetic search

it should be born in mind that the two stochastic update algorithms are much less sensitive to initialisation than their deterministic counterpart.

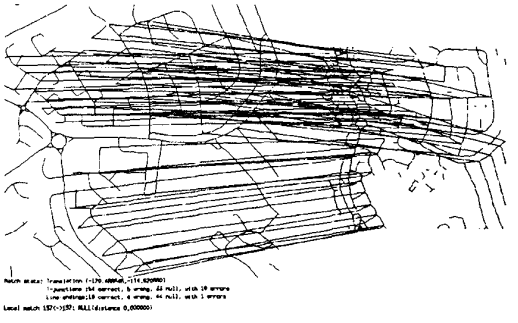

Figure 5: Correct matches

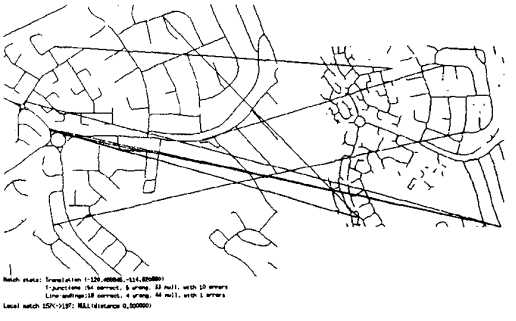

Figure 6: Incorrect matches

6 Conclusions

Our main contribution in this paper has been to explore the use of stochastic optimisation techniques for relational matching. Rather than following the route of applying standard optimisation techniques to essentially *ad hoc* cost functions, we have been a pains to tailor the optimisation process to an objective measure of relational consistency. We show how the apparatus of statistical physics may be applied to this measure to compute configurational Gibbs potentials. With these to hand optimal matches may be located using the Metropolis algorithm. Our second contribution is to demonstrate that because of its Bayesian modelling of relational matching errors, our consistency measure is more naturally suited to optimisation by genetic search. We compare these two stochastic optimisation algorithms with deterministic hill climbing. Here the genetic algorithm offers tangible advantages in terms of rapid and uniform convergence to a global optimum.

References

[1] Aarts E. and Korst J., "Simulated Annealing and Boltzmann Machines", *John Wiley and Sons, New York*, 1989.

[2] Fogel D.B., "An Introduction to Simulated Evolutionary Optimisation", *IEEE Transactions on Neural Networks*, **5**, pp. 3–14, 1994.

[3] Geman S. and D Geman, "Stochastic relaxation. Gibbs distributions and Bayesian restoration of images," *IEEE PAMI*, **PAMI-6** , pp.721–741, 1984.

[4] Hancock, E.R. and J. Kittler, "Discrete Relaxation," *Pattern Recognition*, **23**, pp.711–733, 1990.

[5] Hancock, E.R. and J. Kittler, "A Bayesian Interpretation for the Hopfield Network," *IEEE International Conference on Neural Networks*, pp. 341–346, 1993.

[6] Herault, L., Horaud R., Veillon F. and Niez, J-J., "Symbolic Image Matching by Simulated Annealing", *Proc. BMVC90*, pp. 319–324, 1990.

[7] Peterson C. and Soderberg B., "A New Method for Mapping Optimisation Problems", *International Journal of Neural Systems*, **1**, pp 2-33, 1989.

[8] Shapiro L.G. and Haralick R.M., "A Metric For Comparing Relational Descriptions", *IEEE PAMI*, **7**, pp 90-94, 1985.

[9] Sanfeliu A. and Fu K.S., "A Distance Measure Between Attributed Relational Graphs for Pattern Recognition", *IEEE SMC*, **13**, pp 353–362, 1983.

[10] Wilson R.C and Hancock E.R, "Graph Matching by Discrete Relaxation", *Pattern Recognition in Practice IV, edited by E Gelsema and L Kanal*, pp 165–176, 1994.

[11] Wong A.K.C. and You M., "Entropy and Distance of Random Graphs with Application to Structural Pattern Recognition", *IEEE PAMI*, **7**, pp 599–609, 1985.

[12] Yuille A., "Generalised Deformable Models, Statistical Physics and Matching Problems", *Neural Computation*, **2**, pp. 1-24, 1990.