Inexact Graph Matching
Using Symbolic Constraints

Richard Charles Wilson, B.A.(Oxon)

Submitted for the degree of Doctor of Philosophy

Department of Computer Science

THE UNIVERSITY of York

November 7, 1996

Abstract

Relational graphs are a fundamental type of scene representation for medium and high level computer vision tasks. They provide a generic way of encoding entities and relationships. The comparison and matching of such graphs is an important and challenging problem under the conditions of uncertainty and corruption which exist in most vision problems.

In this thesis a method for matching relational graphs is developed which is based on symbolic constraints. The topology of the graph relations is used to calculate the consistency of a particular match using a Bayesian probability model of the processes at work in the matching process. The result is a global consistency criterion which measures the quality of match. A discrete relaxation technique is used to locate the optimal mapping between graphs using a MAP update rule. The technique is evaluated using both real-world image data and simulated graphs.

Three methods of eliminating spurious elements from the graphs are also studied. The first method involves a constraint-filtering method which is applied after matching has taken place. The second method is an optimisation technique in which noise elements are identified and labelled during the matching process. The final method involves the dynamic reconfiguration of the graphs to remove noise elements during the matching phase. Detailed evaluations of these methods are performed on simulated data.

A theoretical analysis of the criterion is carried out which allows prediction of the expectation value of the criterion at a given level of graph corruption. The performance of the symbolic criterion is compared to that of other alternatives reported in the literature.

Finally the symbolic methods developed earlier are extended both to the use of probabilistic relaxation to match relational graphs, and to the matching of hierarchical graphs.

Contents

List of Figures

List of Tables

Abbreviations

ARG	Attributed Relational Graph
GA	Gradient Ascent
IR	Infra-Red
MAP	Maximum *a posteriori* Probability
PR	Probabilistic Relaxation
SA	Simulated Annealing
SAR	Synthetic Aperture Radar
SPM	Structure Preserving Mapping

Symbols

G	Attributed Relational Graph
V	The set of graph nodes
E	The set of graph edges
$\mathbf{X}$	The set of graph measurements
u, v	Graph nodes
$\mathbf{x}$	Measurement on graph node
C	Graph clique (sub-unit)
$\mathcal{C}$	Size of clique
R	N-ary relation between nodes in graph 1
Γ	Matched realisation of N-ary relation in graph 2
S	Structure Preserving Mapping
$\mathcal{F}$	Consistency Criterion
f	Mapping function
d	Distance function
$\mathcal{P}$	Set of structure preserving mappings
$\mathcal{H}$	Hamming distance
$\mathcal{S}$	Size difference
P_e	Label error probability
P_s	Relational corruption probability
P_ϕ	Node corruption probability

Acknowledgments

Firstly I would like to thank my supervisor, Dr. Edwin Hancock for his support and guidance throughout the course of this work and beyond. My thanks also go to Dr Jim Austin for his assessments of my work.

The DRA, Malvern have provided equipment and data without which this work would have been impossible, and for their financial support through a CASE studentship. In particular I must thank Dr John Radford for his help.

Thanks also go to EPSRC for the funding of my studentship.

On the technical and computing side Andrew Finch and Andrew Cross deserves mention for helping out with many useful programs and ideas.

Also worthy of mention are Andrew Hague, Andy Coombes, Nigel Sharp and Jason Good for their dedicated help with games playing and general time-wasting.

Iain Checkland, Dave Cattrall, Andy Vickers and Simon Moss, along with other sporting heros to numerous to mention, have provided many stress-relieving sporting occasions.

My parents must of course be thanked for all the support they have given me over the years.

Finally above all I would like to thank my fiancée Manique who makes it all worthwhile.

Declaration

I declare that all the work in this thesis is solely my own except where attributed and cited to another author. Some of the material in the following chapters has been previously published in the cited places; for Chapters 2 and 3, (Wilson et al., 1995; Wilson and Hancock, 1994d; Wilson and Hancock, 1994c; Hancock et al., 1994; Wilson et al., 1994); for Chapter 4, (Wilson and Hancock, 1995b; Wilson and Hancock, 1995c; Wilson and Hancock, 1995d; Hancock and Wilson, 1995); for Chapter 7, (Finch et al., 1995; Wilson and Hancock, 1994a; Wilson and Hancock, 1994b; Wilson and Hancock, 1993a; Wilson and Hancock, 1993b) and for Chapter 6, (Wilson and Hancock, 1995a; Cross et al., 1995).

Chapter 1

Introduction and Literature Review

1.1 Introduction

Relational graphs are a fundamental representation of image structure in intermediate and high level computer vision tasks. Their key advantage is the departure from an essentially pixel based representation to a more abstract realisation more closely allied to the intuitive structure of real-world objects. They are therefore a powerful tool for modelling scene structure in terms of objects and the relations which exist between them.

If techniques are available to compare and match these representations, it is possible to interpret scenes in terms of object-based models. This task is not straightforward however; two major problems hinder attempts to match relational graphs. The matching process itself is combinatorially expensive - the number of possible matches rises factorially with the size of the graphs. Furthermore, the relational graph representation of a scene is almost always corrupted by image noise and poor segmentation. In this situation, matching can only be accomplished by inexact means which account for anticipated imperfect matching between graphs.

Successful approaches to this problem make extensive use of measurement information both on objects and on the relations in the graphs to overcome matching ambiguities and graph corruption. This approach relies on apriori knowledge of the parameters of the scene sensing process and the fidelity of the resulting measurements.

In contrast to the measurement-based approach, it is possible to take a symbolic view of

the matching task. Here the graph relations themselves provide the information necessary to effect an accurate match, with the benefit that few parameters are required to perform the matching. Furthermore, the matching process is not heavily reliant on scene measurements. However past approaches of a purely symbolic nature have proved less effective because of the increased ambiguity of relational information. In this thesis we will demonstrate that matching can be accomplished by symbolic means using a suitable model of how topology changes from one graph to another, even when significant corruption is present.

1.2 Review

Relational graph matching has been a task of pivotal importance to intermediate and high level computer vision for some 25 years. It was the work of Barrow and Popplestone (Barrow and Popplestone, 1971) which first exploited a relational graph representation of scene structure. Since then graph-based representations have been a central theme of vision research. As a consequence, a wealth of matching techniques and relational distance measures have been developed. The focus of this thesis and the literature review in the remainder of this chapter is on the matching of relational representations.

One of the first uses of graphs as a tool in computer vision was described by Barrow and Popplestone (Barrow and Popplestone, 1971). In an application involving the matching of a semantic model, relational graphs were used to represent the positional relationships between segmented image regions. The role of the relational graph was as a representation of scene structure which was abstracted away from the low-level pixel-based representations. From these early beginnings the idea of a relational graph to represent scene structure was born. Scene graphs were stored and then images recognised by exhaustive comparison between the graphs. These ideas where later formalised by Barrow and Burstall (Barrow and Burstall, 1976) who presented some practical matching techniques for finding sub-graph isomorphisms based on identifying maximal cliques in the association graph(see below for a discussion of this technique). At this early stage the difficulties of reliably extracting such relationships were not fully appreciated. Infact, these techniques were only effective when the structural scene descriptions were uncorrupted, and therefore had limited practical

application.

It was soon realised that relational models of scene structure could be powerful tools in the computer vision domain. A body of work was already available from mathematical graph theory (Harary, 1969). This dealt with formal definitions of matching between graphs; the problems studied here were graph isomorphism and sub-graph isomorphism. Algorithms have been developed from these foundations to calculate graph and sub-graph isomorphisms (see (Ballard and Brown, 1982) for a review). Of particular interest is the clique problem which is related to sub-graph isomorphism; the objective of the clique algorithm is to find the fully connected sub-graphs or cliques of an input graph. Subgraph isomorphisms between two graphs can be found using this algorithm by forming the association graph of the two graphs. The association graph is formed as follows: consider two graphs G_1 and G_2 with nodes $\{u_i \forall u_i \in V_1\}$ and $\{v_i \forall v_i \in V_2\}$ respectively. The nodes of the association graph A are formed from the node-pairs $\{(u_i, v_j) \forall u_i \in V_1, v_i \in V_2\}$, i.e. there is an association graph node, denoted a_{ij} for each of these node pairings. Edges exist between nodes a_{ij} and a_{kl} when the condition $(u_i, u_k) \in E_1$ and $(v_j, v_l) \in E_2$ is fulfilled. Extraction of the fully connected sub-graphs of the association graph gives the sub-graph isomorphisms between the two graphs G_1 and G_2. These ideas have been exploited more recently by Horaud and Skordas (Horaud and Skordas, 1989) in the context of stereo matching.

A number of important results come out of graph theory concerning the complexity of evaluating graph and sub-graph isomorphisms (Ullman, 1976). The sub-graph isomorphism problem has been found to be *NP-complete*, that is to say that all exact algorithms (those guaranteed to find the correct solution) take a worst case time exponential in the number of graph nodes. Inexact algorithms can solve the problem in polynomial time, but of course without guaranteed success. Full graph isomorphism is *not* known to be *NP-complete*, however no deterministic polynomial time algorithms are known.

In addition to these computational constraints, the problem of noise is also an important factor, and one which is not considered in theoretical and mathematical treatments of graph theory. When graph corruption is a significant factor, which it invariably is in computer vision problems, the perfect isomorphism between graphs is destroyed. Sub-graph isomorphism may still be used to isolate undisrupted portions of the match (Ullman, 1976; Horaud

and Skordas, 1989).

There are, then, two major drawbacks to these these graph-theoretical methods. Firstly the computational complexity is enormous for all but the smallest graphs and rises exponentially with the number of graph nodes. The second problem is their inability to deal with noise, corruption or any form of inexactness within their exact framework. It is clear that if we hope to match noise-corrupted graphs of significant size in an efficient way we must adopt an inexact approach which is robust to graph-errors. It is these techniques which we will examine in the thesis.

1.3 Graph representations of Computer Vision Problems

A scene in a computer vision problem can be thought of as consisting of two distinct elements, a set of objects or tokens from which the scene is constructed and a set of measurements or properties of these objects. The unique element in a relational graph problem is the fact that additional information about the relationships between objects in the scene is available. These relationships, measurements and objects can be used to form what is usually referred to as an *ARG* (*A*ttributed *R*elational *G*raph).

In discussing different representations of an ARG, there are two elements to consider corresponding to the two elements of the graph formulation; what attributes or measurements are used, and how relations between objects are formed.

1.3.1 Graph Relations

The relations between scene objects can be roughly grouped into two separate categories, perceptual relations and abstract relations. Perceptual relationships between objects represent properties of some physical or perceptual significance in the scene; for example adjacency, inclusion, above (Ranganath and Chipman, 1992), corners, junctions and parallelism (Davis, 1979; Etemadi et al., 1991). These are properties which physically exist in the real world and could reasonably be expected to appear in a consistent fashion from image to image. However robust extraction of perceptual object relationships such as these is not an easy task, especially prior to scene interpretation. Moreover, they are rather fragile and

prone to corruption (Sarkar and Boyer, 1993).

Other geometric schemes include pixel-based neighbourhood schemes (Geman and Geman, 1984; Hancock and Kittler, 1990b) and closest point graphs such as the Delaunay graph and Gabriel graph (Tuceryan and Chorzempa, 1991). In (Tuceryan and Chorzempa, 1991), the robustness of various closest point graphs to corruption is investigated. The conclusion reached is that the Delaunay graph is the most resistant to disruption; a result of considerable significance to the matching of noise scenes using abstract relations.

The sensitivity to noise of the perceptual relations is not easy to analyse; the response to noise depends not only on the type of relation but also on the type of objects present in the scene and the segmentation strategy. However (Chipman and Ranganath, 1992) suggest that 'fuzzy' relations with real-valued measurements are more robust and useful than simple binary or "on/off" relations such as 'above' and 'adjacent'.

This idea of associating measurement information with graph relations has been widely adopted (Kittler et al., 1993; Boyer and Kak, 1988). For example the graphs of Kittler *et al* are fully connected - it is the measurements on these relations which provide the distinction between the graph connections. In this sense the relations are 'soft', being characterised by a real-valued measurement. This constrasts with the symbolic approach which is based on the existence or non-existence of relations. This measurement approach clearly provides more information, potentially aiding the matching process. However, as with any measurement process, in order to be of any use the probability distribution of measurements must be defined. This means an increase in apriori information which needs to be known.

Hierarchies

Within the realm of perceptual relations, there has been considerable interest in hierarchical representations of scene structure. The main motivation for this approach is the realisation that the richness in structure of the real world is impossible to model at one level of abstraction. In the hierarchical approach, simple scene structure is first interpreted and this information is used to simplify the analysis of more complex structures. With the hierarchical approach, complex entities can be modelled as simple groupings of less complex structures.

As an example, Dickenson et al (Dickinson et al., 1992) use an "aspect hierarchy" which attempts to closely model the structure of physical objects. They use a hierarchy which builds from contours through contour grouping to faces. These faces can be thought of as possible views of a set of object primitives from which objects themselves are constructed. In this case the important modelling step is to choose a set of primitives from which objects can be constructed; Dickenson *et al* use a rich set of geometric solids (cylinders, cones, cuboids - c.f. Marr's cylinder zoo (Marr, 1984) which constructs objects from cylinders). Clearly it is difficult to choose a set which can represent a wide range of different objects given the variation in structure which real objects exhibit. The need for such a conceptual (or perceptual) hierarchy has been stressed by (Henderson, 1990) in the domain of discrete relaxation.

Another hierarchical approach worthy of mention is the technique of subsumption of detail or resolution. In this approach the top of the hierarchy contains sparse details, information is limited and the resolution is coarse. As we move down the hierarchy scene detail increases, and we can draw on information from the coarser levels to interpret the scene. Examples of such a detail hierarchy can be seen in (Gidas, 1989), (Lau et al., 1993) and (Mjolsness et al., 1989). The pyramidal resolution approach falls within this category (Hancock et al., 1992; Lu and Jain, 1992; Meer et al., 1990), in which different levels represent different image resolutions.

Active Graphs

Most researchers use graphs which are essentially static; that is to say that once formed from scene data, the graphs do not change their structure in the matching phase. Another approach motivated by the need to overcome corruption is to modify the graphs during matching using edit operations (Messmer and Bunke, 1994; Sanfeliu and Fu, 1983). Edit operations such as deleting arcs and nodes are performed until a homomorphism between graphs is found. The number and 'cost' of these operations determines the distance between the graphs under match.

1.4 Graph Matching Criteria

A graph matching criterion is a function of the match between graphs which gauges the quality of the match. If such a criterion is to hand, then the graph matching problem may be accomplished by optimisation. A good choice of criterion or energy function is vital if the optimisation phase is to be fast and accurate and the quality of the final match good. In this section we discuss the range of energy functions present in the literature.

1.4.1 Heuristic Energy Functions

By far the most common way of defining an energy function is in a goal-directed fashion. Terms are added to the energy function to perform specific tasks. These terms are not generally constructed in a rigourous way, rather they are known to have a optimal point at the desired solution and some less favourable value elsewhere, for example (Yuille, 1990). This lack of understanding of the energy function away from the optimal point leaves the way open for problems caused by local optima and slow convergence. Ad-hoc approaches are commonplace in the literature; for instance in probabilistic relaxation approaches, (Ranganath and Chipman, 1992; Izumi et al., 1992; Ton and Jain, 1989) define support functions in an ad-hoc fashion. In the structural domain (Horaud and Skordas, 1989) employ an empirical energy function for choosing between a number of candidate maximal cliques.

It is also possible to impose syntactic constraints on the match such as demanding a one-to-one mapping by the inclusion of additional terms into the energy function. The optimal point is repositioned by the extra term to a point which fulfils the relevant constraint (Yuille, 1990; Mjolsness et al., 1989). However the play-off between the strength of the constraint term and quality of match terms is difficult to control and intervention is often required to set arbitrary weighting factors between the terms. Arbitrary weights are not only confined to the constraint terms and they abound throughout the heuristic approaches (Ranganath and Chipman, 1992; Izumi et al., 1992). Because of the ad-hoc approach, theoretic analysis of the appropriate values for such constants is not possible. Invariably they must be empirically set by the user for different problems.

1.4.2 Continuous Energy Functions

One way to define an energy function in a more principled manner is to use Bayesian or probability-based methods. In this paradigm, a continuous probability of a particular matching is maintained for all matching configurations. Optimisation of this probability leads to the matching configuration with the largest probability. The principle advantage of this approach is that the processes which lead to matching errors can be modelled objectively by probability distributions.

It was (Hummel and Zucker, 1983) who first showed that the probabilistic relaxation scheme of (Rosenfeld et al., 1976) could be interpreted as the minimisation of an energy function based on label probabilities and a set of heuristically defined support functions. Following on from this, (Faugeras and Berthod, 1981) and (Bhanu and Faugeras, 1984) developed similar energy-based criteria which maximised consistency while minimising ambiguity in the labelling. The later work of (Kittler and Hancock, 1989) demonstrated how the process of probabilistic relaxation could be viewed entirely in terms of probability distributions, specifying support functions in terms of these distributions. Consequently the process of PR can now be seen as the optimisation of a consistency function based entirely on probability distributions. Recent efforts by (Kittler et al., 1993) have extended this Bayesian framework to incorporate ARG's with measurement information pertaining to the binary relations within the graph.

Another probabilistic approach is that of (Boyer and Kak, 1988) and (Sengupta and Boyer, 1995). They use an information theoretical approach to define a distance measure between matching units of the graph. This measure which is the conditional information conveyed by the second unit about the first is essentially a measure of entropy between the units. The global graph criterion then becomes the sum of the individual terms for the units, or the global match entropy.

1.4.3 Statistical Physics

An interesting point of contact exists between labelling problems (of which matching problems are a subset) and statistical physics. Statistical physics is concerned with the properties

of a system of large numbers of particles in which the particles can inhabit different energy states. The analogy with labelling problems in computer vision stems from the identification of the particles as objects in the scene and the possible particle states as different possible labellings of the object. Statistical physics tells us that if we can identify a 'potential' U with the system of labels and objects, the probability of the state Ω can be calculated by the Boltzmann distribution:

$$P(\Omega) = \frac{e^{-\beta U(\Omega)}}{Z} \tag{1.1}$$

where $Z = \sum_{\Omega} e^{-\beta U(\Omega)}$ is the partition function and $\beta = 1/kT$ is a constant dependent on the 'temperature' T of the system. Interpreted in the sense of a labelling problem the 'temperature' determines the amount of randomness or uncertainty in the system.

If we adopt this analogy then it gives a direct connection between the definition of an energy function and the probabilities of various system states. It is therefore possible to analyse any defined energy function in terms of probabilities. Geman and Geman(Geman and Geman, 1984) applied these concepts with seminal impact to low-level vision and used them to develop a non-deterministic updating scheme referred to as simulated annealing. This scheme uses the temperature as a control variable which is used to introduce controlled amounts of randomness into the updating schedule. Geman and Geman were able to show the equivalence between the Gibbs distribution and a Markov Random Field for calculating the MAP estimate of the labelling. The MAP estimate of the probability distribution defined in equation 1.1 is recovered by the process while local minima are avoided. This method is discussed further in section 1.5.1

The relationship between minimising an energy function and a probabilistic approach to the matching problem is explored further by Yuille (Yuille, 1990). He reiterated that any problem formulated in terms of an energy function can be given an interpretation in terms of probability by the Gibbs distribution. Yuille then uses the partition function to map hard constraints onto an energy function.

1.4.4 Structural Pattern Recognition

As described above, the ARG under match contains a considerable amount of measurement information. We can loosely group matching schemes into two categories based on the way in which they utilise the information. Those that only use measurement information in an initialisation phase, after which labelling decisions are made by a purely symbolic process are referred to as 'symbolic schemes'. Those in which the measurement information persists and is referred to at all stages are 'measurement-based schemes'.

When designing a matching criterion the question arises of how much of this information should be used to aid the matching. At first sight the answer is simple - the more information available, the better the result will be. Indeed much recent work has adopted this philosophy with reasonable success. After earlier work which focussed mainly on symbolic processes (for example (Shapiro and Haralick, 1985; Horaud and Skordas, 1989)) the need was seen for increasing amounts of scene information to disambiguate the scene interpretation process. Classical structural pattern recognition was perceived as impoverished in terms of the amount of scene information available. However this is not necessarily the case as there are both advantages and disadvantages to a heavy reliance on scene measurements. On the positive side they decrease the amount of ambiguity in the scene interpretation phase and alleviate the need for potentially fragile structural information. On the other hand, parameters of the measurement probability distributions must be known before hand and the matching can be inflexible to differing scene conditions. This inflexibility is due to parameterisation of the probability distributions. The parameters often vary from scene to scene and consequently need to be measured and adjusted for each new matching task. On the other hand, symbolic matching approaches can overcome some of the problems of ambiguity and lack of reliable measurement information by successfully using contextual information.

Proponents of the measurement approach include (Boyer and Kak, 1988) with their conditional mutual information based on both inter-object and inter-relation measurements. Kittler et al (Kittler et al., 1993) use unary measurements on nodes and binary measurements between nodes in their probabilistic relaxation scheme. Mjolsness et al (Mjolsness et al., 1989) use inter-object measurements in their Hopfield-style energy function, which

is optimised with a neural network. Ranganath and Chipman (Ranganath and Chipman, 1992) similarly use inter-object measurements in the definition of a support function for probabilistic relaxation, albeit in an ad-hoc fashion.

Symbolic approaches include the original relaxation scheme of (Rosenfeld et al., 1976); their definition of a support function and relaxation scheme is such that measurements were used in the initialisation of probabilities, but as the scheme progresses this information is overridden by labelling constraints. This has been seen as a weakness of the classic PR technique. Recent work on relaxation has focussed on creating a persistence of measurements in the PR technique by incorporating measurement probability distributions into the support function (Kittler et al., 1993)

Part of the aim of this thesis is to demonstrate that symbolic techniques can overcome the problems of ambiguity by effectively using contextual information and gain the advantages of flexibility, robustness and ease of control.

1.5 Optimisation

When a matching criterion is available for the problem under study, which has, for the sake of discussion, a maximum point at the desired solution of the problem, the task is then to locate this maximum point by some optimisation technique. This is not in general a simple task; as we discussed earlier the space of possible mappings between a pair of graphs under match is enormous. For example two graphs each of only 10 nodes each has 10! possible matches (4×10^6 combinations). Clearly for graphs of moderate size direct evaluation of all matching combinations is not feasible. Rather than take this brute force approach, all authors have employed methods of traversing the search space from a poor initial position to the optimal point, using only a few intermediate steps. It is these methods to which we now turn our attention. It is worth pointing out at this stage that traversing the search space is not in general a simple task; if local optima exist which do not correspond to the global optimum the potential exists for incorrect solutions to be found. Furthermore the task becomes increasingly difficult when the initial guess is far from the final target.

Optimisation schemes may be split broadly into two categories; continuous optimisation

methods operate when a continuous labelling space is available. These methods are well understood mathematically (Faugeras and Berthod, 1981). In contrast discrete optimisation methods in which the mapping is discretized in labelling space is not well understood.

1.5.1 Continuous Optimisation

We consider first the case when the matching criterion is continuous, i.e. the case when the space of labellings is continuous and the matching criterion exists at all points. In more formal notation, we have a real-valued mapping function which tells us how 'likely' the match from node u in G_1 to node v in G_2 which we will refer to as $\rho(u \rightarrow v)$. This function always resides in the interval $[0, 1]$. The components $\rho(u \rightarrow v)$ can be formed into a vector of matching weights denoted by $\vec{\rho}$. The matching criterion $\mathcal{F}(\vec{\rho})$ is defined for all $\vec{\rho}$ and is therefore continuous.

Gradient Ascent

Imagine that we are interested in the maximum point of the matching criterion $\mathcal{F}$. If $\mathcal{F}$ has just one maximum and monotonically increases to this maximum the problem is relatively straightforward. We need only move through the space of mappings in the direction of the maximum gradient of $\mathcal{F}$ to locate the optimal point. Formally

$$\Delta\vec{\rho} = \nabla\mathcal{F}(\vec{\rho}) \tag{1.2}$$

This technique is known as gradient ascent (GA).

Projected Gradient

The GA technique above is sufficient to locate the maximum point of $\mathcal{F}$ when there is one maximum and $\mathcal{F}$ monotonically increases to this maximum, and, importantly, when no constraints apply to the components of the mapping vector $\vec{\rho}$. We referred earlier to the components of $\vec{\rho}$ as 'likelihoods' because they do not necessarily represent probabilities. However many authors wish to interpret them as probabilities and for this to be the case the components must obey the axiomatic rules of probability. In this case they must be

subjected to the constraint

$$\sum_{v \in V_2} \rho(u \rightarrow v) = 1 \tag{1.3}$$

which implies that each node in G_1 should have only one match in G_2. With this constraint $\rho(u \rightarrow v)$ may be interpreted as the probability that node u is matched to node v.

If Equation 1.3 is to satisfied we can no longer simply update the matching-vector $\vec{\rho}$ by the gradient of the matching criterion since we have no guarantee that this will lead to an allowed value of the matching-vector (see (Bhanu and Faugeras, 1984; Zucker and Mohammed, 1978; Faugeras and Berthod, 1981; Luo et al., 1988)). Instead we must use a gradient projection method. In this method the constraints in Eqn. 1.3 are viewed as defining a sub-space in the labelling space which contains only the allowed values of $\vec{\rho}$. We then project the gradient vector (Faugeras and Berthod, 1981) onto this sub-space to obtain an update vector who's magnitude and direction are determined by the gradient but keeps the mapping-vector in probability space. If the projection operator is denoted $\mathcal{P}$, then the update rule becomes

$$\Delta\vec{\rho} = \mathcal{P}\nabla\mathcal{F}(\rho) \tag{1.4}$$

Non-unit step sizes

While these gradient ascent techniques will converge to the optimal value under the conditions we have specified, the rate of convergence may not be satisfactory. The simple approach we have so far examined always takes a unit step towards the maximum, which may take considerable time if the gradient is small and the maximum is distant. Under these conditions we would like to take a larger step. When the maximum is near by we wish to take smaller steps in order to avoid overshooting the optimal point. This approach has been explored by (Faugeras and Berthod, 1981; Lloyd, 1983) and (Bhanu and Faugeras, 1984). The update is defined by (in the case of the gradient projection method)

$$\Delta\vec{\rho} = \alpha\mathcal{P}\nabla\mathcal{F}(\vec{\rho}) \tag{1.5}$$

where α is the step size of the update. It then remains to determine the appropriate value for this step. Faugeras and Berthod (Faugeras and Berthod, 1981) show how this can be achieved analytically when the energy function is quadratic; in this case the ideal step size

is easily determined. For polynomial energy functions of high order and other non-linear functions, the step size is not easy to calculate; usually the local properties of the function are interpolated by a quadratic or cubic polynomial and the corresponding step size is only approximate.

Probabilistic Relaxation Schemes

Relaxation schemes are optimisation techniques in which the variables of the scheme are iteratively updated in order to approach a stationary point of the update equations. They can be used to optimise a matching criterion which has a maximum at the stationary point. The exciting feature of relaxation algorithms for matching problems is their inherently parallel nature.

The classic probabilistic relaxation scheme is due to Rosenfeld, Hummel and Zucker (Rosenfeld et al., 1976), and was conceived as an object labelling algorithm. Since it's conception the approach has been widely used for image processing tasks including graph matching.

Again the matches are represented by a matching vector, but this time more in the spirit of probabilities; we will denote the probability that u matches to v at iteration s of the scheme as $P^{(s)}(u \rightarrow v)$. The Rosenfeld *et al* scheme specifies that the probabilities at iteration $s+1$ should be given by

$$P^{(s+1)}(u \rightarrow v) = \frac{P^{(s)}(u \rightarrow v)Q^{(s)}(u \rightarrow v)}{\sum_{w \in V_2} P^{(s)}(u \rightarrow w)Q^{(s)}(u \rightarrow w)} \tag{1.6}$$

where Q is called the support function. The reason for choosing this particular form becomes clearer when we study the individual terms. The numerator appearing in the RHZ update formula can be viewed as the product of two 'probabilities' or evidential factors; the first being the probability that node u is matched to v, and the second being the probability of the surrounding matches given that v is the match on u. This then combines the local probability with contextual information from the surroundings. The denominator simply ensures the normalisation of the probabilities, i.e. that $\sum_v P^{(s+1)}(u \rightarrow v) = 1$. This step is necessary since the support functions Q are not true probabilities.

The scheme depends on the definition of a support function. Rosenfeld *et al* define the support function as

$$Q(u \rightarrow v) = \sum_{n \in V_1} C_{un} \sum_{m \in V_2} r(u \rightarrow v, n \rightarrow m) P(n \rightarrow m) \tag{1.7}$$

where C_{un} are arbitrary weights specifying varying influences of neighbouring nodes n on u, and r determines how compatible the matches $u \rightarrow v$ and $n \rightarrow m$ are.

In the original form there are several problems with this scheme. The support function, compatibility coefficients and weights are purely arbitrary - no method for their specification is offered. No convergence properties are offered; it is not clear if the scheme will converge at all. Finally no meaning in terms of an energy function is attached to the stationary points of the scheme. Hummel and Zucker (Hummel and Zucker, 1983) later overcame some of these problems, showing that the RHZ scheme infact optimised the energy function

$$\mathcal{F}_p = \sum_{u \in V_1} \sum_{v \in V_2} P^{(s)}(u \rightarrow v) Q^{(s)}(u \rightarrow v) \tag{1.8}$$

and at the stationary points, the probabilities were unambiguous, i.e. $P_s(u \rightarrow v) = 0 \; or \; 1$.

Despite drawbacks probabilistic relaxation proved to be a powerful tool in labelling problems. A plethora of alternatives evolved, and due to a lack of theory behind the specification of support functions they have many different and generally heuristic definitions of support. See (Price, 1985) for a comparison of some of the more conventional approaches. This final hurdle was recently overcome by Kittler and Hancock (Kittler and Hancock, 1989) who show how the probabilistic relaxation algorithm can be see as the iterative filtering of a set of object measurements. The filtering is based on the maximisation of the conditional probability of the object labels given the measurements i.e. $P(u \rightarrow v | \mathbf{x}_w \forall w \in V_1)$. When this Bayesian approach is adopted, the resulting scheme includes a specification of the support function in terms of probability distributions:

$$Q^{(n)}(u \rightarrow v) = \frac{1}{P(u \rightarrow v)} \sum_{\theta \in \Omega} \Big\{ \prod_{n \in C_u} \frac{P^{(n)}[n \rightarrow \theta(n)]}{P[n \rightarrow \theta(n)]} \Big\} P[\theta(m) \forall m \in C_u] \tag{1.9}$$

Where C_u is the neighbourhood of u, that is u and all its interacting neighbours. θ represents the set of possible matches on C_u and Ω is the set of possible labellings on these nodes. This support function is however of exponential complexity which limits it's usefulness in realistic labelling problems.

Kittler and Hancock provide two methods for reducing the complexity of this expression. The first method involves defining a limited dictionary of allowed labellings on any graph neighbourhood. If this dictionary of labellings is denoted Θ, then the expression for support is given by

$$Q = \frac{1}{P(u \rightarrow v)} \sum_{\theta \in \Theta} \Big\{ \prod_{n \in C_u} \frac{P^{(n)}[n \rightarrow \theta(n)]}{P[n \rightarrow \theta(n)]} \Big\} P(\theta) \tag{1.10}$$

The sum is over the limited set of dictionary items rather than the entire space of mappings.

The second technique involves factorisation of the support function; details of this approach are discussed in Chapter 7; suffice it to say here that under certain independence assumptions the support function can be simplified to one of polynomial complexity.

In a more recent development (Kittler et al., 1993) have developed this methodology to incorporate binary measurements between pairs of objects; They are interested in evaluating $P(u \rightarrow v | \mathbf{x}_w \forall w \in V_1, A_{ij} \forall i, j \in V_1)$.

1.5.2 Discrete Optimisation Techniques

Discrete Relaxation

Discrete relaxation differs from probabilistic relaxation in that there is no ambiguity in the labelling at any point during the scheme - at all times a match is maintained. This technique pre-dates probabilistic relaxation with the discrete labelling ideas of (Waltz, 1975), which evolved from earlier work on consistent labelling problems by (Huffman, 1971; Clowes, 1971). Indeed the probabilistic version was a development of Waltz's discrete method. The simple idea of this approach is to visit each object in turn and update the label on that object in order to gain the maximum improvement in the matching criterion of the problem. The update rule is therefore

$$f^{(n+1)}(u) = \arg \max_{f^{(n)}(u) \in V_2} \mathcal{F}[f^{(n)}(u), f^{(n)}(v) \forall v \in V_1, v \neq u] \tag{1.11}$$

More recently the discrete relaxation labelling problem has been cast into a Bayesian probabilistic framework (Hancock and Kittler, 1990a). They suggest that the matching criterion should be defined as the joint probability of the labelling given the (unary) measurement

information, i.e. $P(f|\mathbf{X})$. Under assumptions of independence of the unary measurements they show how the discrete update procedure can be used to locate the MAP estimate of the labelling.

Simulated Annealing

In their seminal paper, Geman and Geman developed an interesting variant on the basic relaxation scheme (Geman and Geman, 1984). Drawing on analogies with statistical physics, they show how a set of discrete labels on objects, if placed in a Markov Random Field, can be viewed as a system of interacting particles. A Gibbs distribution can then be associated with the labels. The interesting feature of the Gibbs distribution is the rôle of a 'temperature' variable; this variable has the effect of introducing controlled levels of smoothing into the energy function. They then coupled this energy function with a semi-random label update procedure in which updates that increase the energy function are also allowed with a certain probability, as well as those which decrease the energy function. This allowed the optimisation phase to escape small local minima in the energy function. They were also able to show that if temperature was reduced according to a certain rather restrictive schedule, then all local minima would be avoided and a global optima would be found.

The main draw-back of their method is the speed of convergence. The proposed reduction schedule is impractical and the update procedure computationally expensive.

More recently (Herault et al., 1990) have adopted the simulated annealing technique for performing structural matching tasks.

1.5.3 Search Techniques

Search techniques involve sifting through the different configurations of the labelling space in order to locate the optimal point of the energy function. As we mentioned earlier the space of possible mappings is large enough to preclude the possibility of searching all configurations and therefore the key element of search algorithms are their ability to intelligently prune the search space. Examples of search techniques applied to graph matching problems can be seen in (Shapiro and Haralick, 1981) and (Dickenson et al., 1992). These algorithms rely on heuristics and thresholding in order to reduce the space of mappings which need to

be explored (see (Ballard and Brown, 1982) for a comprehensive review).

1.6 Summary

In summary, we should look towards inexact optimisation techniques if we wish to match large, noise corrupted graphs. These provide the necessary robustness to corruption and error while being relatively efficient at locating the best match. On the optimisation side, relaxation techniques provide an attractive method because of their parallel nature.

Study of the different approaches to defining an energy function (either implicit or explicit) reveals an interesting split between the structural and symbolic approaches ((Horaud and Skordas, 1989; Shapiro and Haralick, 1985) for example) and the measurement-based methods (Boyer and Kak, 1988; Kittler et al., 1993). We take the view that structural approaches can be more flexible under varying scene conditions and that it is an advantage to have as few parameters as possible to be specified in advance. In the past symbolic approaches have proved ineffective due to insufficient scene information being available. However if a measure of structural similarity is sufficiently fine, enough information should be available to interpret the scene. We will therefore concentrate on the symbolic matching paradigm.

In order to provide a rigorous framework for developing our models, the appropriate choice seems to be Bayesian probability. Adopting this framework means we are able to model corruption and noise processes with objective probability distributions.

Chapter 2

Comparing Relational Graphs

2.1 Introduction

In this chapter we derive a measure of the quality of a graph match, based on structural considerations and developed using the principles of Bayesian probability theory. In the first half of the chapter we set up a formal framework for graph matching problems and discuss the application of probability theory to such problems. In the second part we develop a new criterion based on structural, topological constraints.

2.2 A Graph-Matching Formalism

The first step in discussing a relational graph matching technique is to establish a suitable formalism for describing the matching process. Here we describe the attributed relational graph (ARG) widely used in the literature (Barrow and Popplestone, 1971; Tang and Lee, 1992; Kittler et al., 1993; Ton and Jain, 1989). Our particular formulation is relatively simple; structural considerations are our primary concern here.

A relational graph is represented by the triple $G = (V, E, \mathbf{X})$ and consists of a set of nodes $V = \{v_1, v_2, ..., v_n\}$ which represent objects in a scene graph or model. The set $E = \{e_1, e_2, ..., e_m\}$ is the set of graph edges, and these represent the presence of a relationship of some sort between a pair of objects. If the objects represented by v_1 and v_2 are related in the image, then they are connected by an edge $e = (v_1, v_2) \in E$. The edges could represent a perceptual relation such as adjacency or Voronoi neighbours. Within our approach the

significant element is the existence of a connection and different types of edge are not distinguished in the formalism.

We also hypothesize that a set of unary measurements are available on the objects in the scene, and these are denoted by the set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$. In the text we simply refer to the measurement for node u as $\mathbf{x}_u$. Our interest will be confined mainly to methods of gathering structural information from the graphs and therefore measurement information is not our primary concern. It is for this reason we have adopted a simple model of measurement information. In this formulation we assume that only unary measurement information is relevant to the labelling of the nodes. While this is a common approach (Barrow and Popplestone, 1971; Rosenfeld et al., 1976; Hancock and Kittler, 1990a) it is far from the only technique. For example (Boyer and Kak, 1988; Kittler et al., 1993) use measurements defined on the relations as an additional source of information.

Our aim in matching is to associate the nodes in a graph $G_1 = (V_1, E_1)$ with those in a second graph $G_2 = (V_2, E_2)$. Graph G_1 is referred to as the data graph and G_2 as the model graph. Nodes from G_1 are denoted by $v^{(1)}$ and those from G_2 as $v^{(2)}$. In order to discuss matching problems we require a mapping function from the nodes in V_1 onto those in V_2:

$$f : V_1 \rightarrow V_2$$

Hence $(u^{(1)}, v^{(2)}) \in f$ denotes the match of node $u^{(1)} \in V_1$ against node $v^{(2)} \in V_2$. There is no explicit restriction on multiple matches; the structural constraints governing our matching process encourage unambiguous matches implicitly. The function f is potentially many to one and therefore non-invertible.

2.3 Matching: The Bayesian approach

Many approaches to matching suffer from one major flaw: the quantitative criterion of matching quality is arrived at by largely goal-directed methods (Ranganath and Chipman, 1992; Izumi et al., 1992; Ton and Jain, 1989; Sanfeliu and Fu, 1983) Therefore, while the matching technique may be effective, there is little meaning in terms of the quality of the match associated with the value of the criterion and theoretical analysis of the algorithm is

intractable.

In order to tackle this problem, an objective framework for quantifying matching performance is required. Bayesian probability theory provides us with just such a framework. The Bayesian minimum error decision rule (Kittler and Taylor, 1994) specifies how best to assign labels to objects based on the label probability distributions and the costs of the different mis-labellings. In the context of our study, the labels of interest are the various matching assignments and the Bayes decision rule tells us how to best assign the matches to minimize the error due to misclassification. Furthermore the framework specifies how to combine probabilities and measurement distributions in a principled way.

Using this framework, we can determine the best set of matching assignments given the information provided by the available measurements by finding the maximum *a posteriori* (MAP) probability of the matching function given the measurements. In other words, the matching criterion is given by the *a posteriori* probability of the match thus:

$$\mathcal{F}(f) = P(f|\mathbf{X}) \tag{2.1}$$

and we should attempt to maximize this quantity with respect to the matching function f.

Two distinct sources of information are at our disposal when studying the matching problem. Observational information is originally provided by the sensor or sensors which image the scene. From the symbolic perspective this data is then processed by a segmentation algorithm to extract individual scene objects and corresponding attributes. Since these attributes reflect measured data about the world, we expect a degree of variability or uncertainty to be present. In order to capture this variability in the measurement information the models are specified in terms of the conditional probability densities $p(\mathbf{x}_i|f(v_i^{(1)}) = v_j^{(2)})$.

The second source of information is present in the structure of the graphs under study. The graph connections provide a source of contextual information which can be exploited to gauge the quality of match. These relationships constrain the matches, since we would expect the relationships between objects to be the same in both data and model scene. This expectation creates a limitation on the set of legitimate graph to graph matches. This is the motivation behind subgraph isomorphism approaches (Ballard and Brown, 1982). However

since we wish to operate under conditions of noise, we must soften these expectations and use inexact constraints.

In the formulation of Equation 2.1 there is no clear rôle for either source of information; the various sources of information are hidden within a single term. By applying Bayes theorem to Equation 2.1 the individual ingredients become clearer

$$\mathcal{F}(f) = \frac{p(\mathbf{X}|f)P(f)}{p(\mathbf{X})} \tag{2.2}$$

A clear dichotomy between the rôle of measurements and structure is now evident. Measurement information is confined to the measurement density $p(\mathbf{X})$ which is not dependent on the match, and to a conditional probability density $p(\mathbf{X}|f)$ which models the probability of the known measurements given a match f. Such models are not our primary concern here and consequently we adopt a simple and widely held assumption (Kittler and Hancock, 1989; Hancock and Kittler, 1990a) that measurements on nodes are conditionally independent of each other. As a consequence we can factorise the joint conditional probability thus

$$p(\mathbf{X}|f) = \prod_{v_i \in V_1} p(x_i|f(v_i)) \tag{2.3}$$

vastly simplifying the calculation of the influence of the measurements.

We should also note at this point that the joint measurement density $p(\mathbf{X})$ does not change during the matching process and for our purposes can safely be ignored.

The second element is then the joint prior $P(f)$. In contrast to the other term, this models the structural aspects of the graphs under match. This knowledge is captured in terms of an *a priori* model of the allowable configurations of matching labels which maps one graph on to the other. The fundamental assumption here is that relational structure should be preserved under the mapping in question. It is exactly how to formulate the input of structural information into the matching process that we intend to study in depth in this thesis.

2.4 Matching Complexity

To match two graphs the simplest approach we can take is to use a non-contextual method. As we mentioned earlier, the joint prior $P(f)$ models the rôle of contextual information in the matching process. The non-contextual limit simply corresponds to the case when all matching configurations are equally likely and no contextual information is present. Under these circumstances $P(f)$ simply has a uniform distribution. Combining Equations 2.2 and 2.3, our task is to maximize $\mathcal{F}_{nc} = \prod_{v_i \in V_1} p(x_i|f(v_i))$. We also note that under our assumptions the node measurements are independent of each other and we may individually maximize each term to find the globally optimum labelling. We therefore compare the unary measurements on the nodes thus

$$f(u_i^{(1)}) = \arg \max_{v_k^{(2)} \in V_2} p(\mathbf{x}_i|f(u_i^{(1)}) = v_k^{(2)}) \tag{2.4}$$

Typically modelling of the p.d.f. is based on a monotonic decreasing function of some distance measure $d(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)})$ between measurements on the two nodes. Example distributions include $p[\mathbf{x}_i|f(u_i^{(1)}) = v_j^{(2)}] = \exp[-d(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)})]$ or $1/[1 + d(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)})]$. A variety of distance measures are also used, including the Mahalanobis distance and the Euclidian distance.

In more complex and realistic pattern recognition problems the unary measurements $\mathbf{X}$ rarely provide sufficient information to allow an accurate match. In this situation an additional supply of information is needed to label each object, and one source of this information can be provided by the object's context within the surrounding objects. The relationship between objects within the graphs provide constraints which we can exploit to enhance the quality of the match. These constraints must be satisfied by a correct match.

With this knowledge, another approach suggests itself to us. We need only locate a mapping in which the constraints are fully satisfied. In our formulation of the problem this involves a binary distribution for the configuration probability. In other words $P(f) = 1$ if f represents a mapping in which all constraints are fully satisfied and $P(f) = 0$ in all other situations. By rejecting any partial match which is inconsistent, the space of mappings which need to be explored remains small and the ideal match can be quickly located. This

approach is at the heart of graph search algorithms (Flynn and Jain, 1991; Jones and Wong,) such as subgraph isomorphism and maximal clique finding (Messmer and Bunke, 1994; Horaud and Skordas, 1989; Horaud et al., 1990; Barrow and Burstall, 1976; Herault et al., 1990); structure which is inconsistent is immediately rejected.

This model however fails to take account of one of the fundamental problems of image processing - data extracted from images is invariably corrupt and uncertain. Because this is the case, we must admit the possibility of both extraneous and missing objects, and constraints which are also corrupt. As an immediate consequence we can no longer discard inconsistent matches as incorrect since they could be the result of graph corruption. Indeed, a fully consistent match in all likelihood no longer exists. Furthermore this hard probability model provides a very coarse measure of the consistency of a match. If we were to use this approach in an optimisation scheme, it would lead to problems in determining the optimal update direction necessary to move towards a more consistent match. In this situation the matching algorithm will often become deadlocked in an inconsistent configuration. Under conditions of corruption the need is for a measure of consistency which uses softer, inexact, constraints and produces a fine measure of consistency.

A finer way of gauging consistency is to count consistent edges in the graph match. However this has the disadvantage of weakening the constraints provided by graph structure and does not make full use of the available structural information. For example, a set of consistent yet incorrectly ordered edges would give a high consistency measure. Another related approach is that of Shapiro and Haralick (Shapiro and Haralick, 1985) which counts consistent cliques. While this uses more structural information, the measure of consistency is still coarse.

For our purposes the approaches presented in the literature are unsuitable; we require a consistency measure which makes powerful use of structural information while providing a fine measure of consistency. This structure defines the relationships between scene objects and therefore controls the flow of contextual information. The symbolic approach has been largely ignored in recent work, in favour of attribute based schemes which use more measurement information and less structural constraints; it was felt that structural methods were too ambiguous to allow the matching of scene graphs without the need for considerable

attribute information. Here we try to show that a fine measure of structural similarity is sufficient to allow matching, and therefore we anticipate a strong role for contextual information and a correspondingly weak influence from measurement information. The algorithms in the literature take the opposite view; Boyer and Kak (Boyer and Kak, 1988) base their matching process on joint measurement information between pairs of objects in the graph as do (Kittler et al., 1993) in their probabilistic relaxation scheme. Such probabilistic relaxation techniques which are based on the quadratic support function (Rosenfeld et al., 1976) may also be seen as adopting an approximation of weak context (Kittler and Hancock, 1989) which is unsuitable when the context provides the major source of information.

2.5 A Structural Approach to Matching Graphs

In summary, a number of problems present themselves. Since the graphs may be corrupt, we cannot eliminate any matching configurations as illegitimate and we must potentially explore them all. However the number of possible matching combinations is $n!/(n-m)!$ where n is the number of nodes in the larger of the two graphs and m the smaller. This number rapidly becomes unmanageable, growing exponentially with n for large n, making direct exploration impossible.

One approach to this problem is to break the large graph down into manageable sub-units (Faugeras, 1981). These sub-graphs can be seen as representative of the graph as a whole. These units can then be matched exactly by a full exploration of the set of mappings between them, provided they are small in size. We can then allow these sub-units to interact through the passing of contextual information between them. This allows the constraint information to propagate across the graph. This is done by allowing the units to overlap, so that two adjacent units contain information about each other via the mutual graph nodes they contain. By applying a suitable relaxation process, the matches can be modified until they best fulfil the constraints applied to them. Because neighbouring cliques overlap, there is effectively a chain of constraints across the entire graph and hence an influence from all matches on all others, provided we iterate the match update procedure a sufficient number of times.

To begin the construction of a global consistency measure based on this principle, we must select graph sub-units appropriate for the task. Units of arbitrary size may be used, however the unit size plays a key role in determining the effectiveness the scheme.There are two issues at work in selecting structures appropriate to this task. One key element is the relative size difference between the set of possible labelling combinations and the set of allowable mappings of the nodes, i.e. the reduction of the search space achieved by applying the topological constraints. As we discussed earlier (section 2.4) the number of possible matches between graphs is $n!/(n-m)!$ if all matches between graphs are allowed. Of course this equally well applies to the sub-graphs we are trying to match here. However if we use the sub-graph relational structure by only allowing matches between units which have the same topology, we can drastically reduce the space of legitimate mappings. It is through this topological constraint that contextual information is introduced into the matching scheme. The greater the reduction in the legitimate space of mappings, the more powerful the applied constraints are.

Of course the topological constraints discussed here are not completely valid under conditions of graph corruption, and therefore an inexact means of gauging the quality of match is required if we are to successfully apply them. By modelling the noise processes which corrupt the graph we can still apply the structural constraints in an error tolerant form.

The more constraining the topological structure is, the more effective the scheme is in discarding unacceptable labellings. In these terms small structural units perform badly and the matching process is impoverished in terms of the contextual information upon which it can draw in locating a consistent match. This limits the effectiveness of the relaxation scheme, rendering it susceptible to noise or error. If, on the other hand, the structural units are too large, then the matching process becomes excessively cumbersome in terms of its computational requirements; the limitation stems from the need to explore the space of mappings between representational subunits.

As an example, consider the case of a pair of joined nodes and a triplet in an example graph (Figure 2.1). There are 30 possible matches of the two nodes onto the graph when we do not consider topological constraints, and 10 possible matches of the pair unit as a whole.

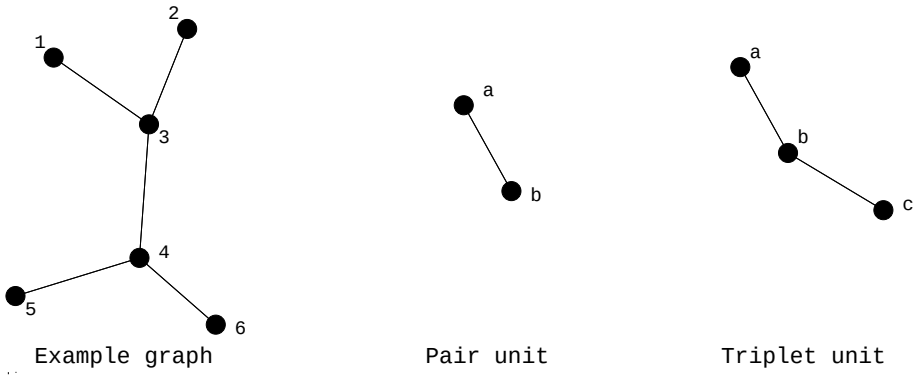


Figure 2.1: Example mappings of different sized units

For the triplet, there are 120 possible matches of the nodes, but still only 12 matches for the unit. While we must explore more combinations for the triplet, the constraining power of the relation is greater as it reduces the number of possible mappings by a factor of 10 rather than 3. When larger units are used, there is a greater reduction factor, but the actual number of mappings still increases.

As a compromise between representational power and computational requirements we propose the choice of sub-graphs which consist of a central node and all the adjacent nodes connected to it by a graph edge. It is important to stress however that the methodology presented here is not limited to a particular sub-graph unit, but is applicable to any type of structural unit. For convenience we will refer to these structural units or N-ary relations as cliques.

2.6 A Graph Matching Criterion

2.6.1 Structure preserving mappings

As mentioned above we have adopted the clique as our basic structural unit, which is denoted by C. This consists of a central node and all those linked to it by a graph edge. More formally, the clique of the node indexed j in the graph G_1 is given by the set of nodes $C_j^{(1)} = j \cup \{i | (i,j) \in E_1\}$. For notational ease, we will denote this N-ary relation as

$$R_j = (u_1, u_2, \ldots, u_{|C_j^{(1)}|}) \tag{2.5}$$

where u_1 is the centre node of the clique. The matched realisation of this relation is therefore given by

$$\Gamma_j = (f(u_1), f(u_2), \ldots, f(u_{|C_j^{(1)}|})) \tag{2.6}$$

That is to say, Γ_j is the set of nodes which the relation R_j currently maps onto. The mapped unit Γ gives us a set of nodes which we can compare with similar relational units in graph G_2 to gauge the quality of the match. The structural unit $C_j^{(1)}$ must match to a similar unit $C_k^{(2)}$ in graph G_2. It is through exploiting this knowledge that G_2 provides constraints on the matching. Since we have no apriori knowledge of the match, any of the possible units generated from G_2 provide a feasible match for $C_j^{(1)}$. Furthermore we do not know how the nodes from the relation R_j map onto the nodes of any candidate clique from G_2. We need to perform a full graph-to-graph matching between these sub-graph units by exploring all possible mappings between them which preserve the topological structure of the relations. In practice, if the candidate clique from G_2 is $C_k^{(2)}$ we permute the set of nodes in $C_k^{(2)}$ through all combinations which preserve the adjacency structure of the sub-graph unit, to form a set of M potential mappings of $C_k^{(2)}$; Each potential mapping is denoted by

$$S_k^m = (v_1, v_2, \ldots, v_{|C_k^{(2)}|}) \tag{2.7}$$

which henceforth will be referred to as a structure-preserving mapping (SPM). The clique $C_k^{(2)}$ generates M possible mappings onto which R_j may legitimately map. The set of structure preserving mappings at node k is therefore given by $S_k = \{S_k^1, S_k^2, \ldots, S_k^M\}$. We can now construct a set $\mathcal{P}(C_j^{(1)}) = \{S_i | \forall i \in V_2\}$ containing all possible mappings of the clique $C_j^{(1)}$. This is a union of all the possible mappings at each clique in G_2. The set of mappings $\mathcal{P}$ contains all legitimate mappings of a clique onto graph G_2. As a consequence the set of mappings $\mathcal{P}$ is identical for all cliques, i.e. $\mathcal{P}(C_j^{(1)}) = \mathcal{P}(C_i^{(1)}) = \mathcal{P}$.

An example set of SPMs is shown in Figure 2.2. In this case the clique consists of a central node and three neighbours. When matching a clique from G_1 onto a clique in G_2, the centre nodes are clearly identifiable and must match to each other. This is not true of the external nodes however, and for these any combinations of matches amongst themselves are

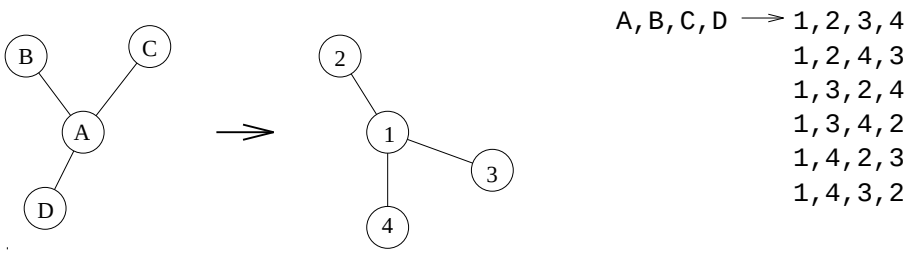


Figure 2.2: Example clique mapping

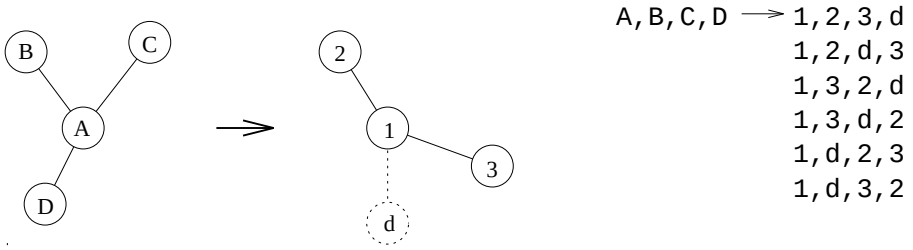


Figure 2.3: Example clique mapping with dummy nodes

legitimate because the topology is preserved. We must therefore run through all possible permutations of the external nodes.

When the cliques are of different sizes, the situation is more complicated (Figure 2.3). We must hypothesize that an unknown number of nodes have been lost or added in one or both of the cliques which disrupt the structure. A model must be adopted at this point to account for this discrepancy; dummy nodes may be added to either clique to account for the missing or extra nodes. These nodes are not added freely, in contrast to (Shapiro and Haralick, 1985)) they are subject to a penalty which is discussed later. In theory since there is no knowledge regarding the number of extra nodes present, we should run through all possible numbers of added dummy nodes to either clique in order to calculate the matching quality. In practice, however, the addition of dummy nodes can be penalised in such a way as to make the mappings with the minimum number of added nodes the dominant term in the calculation of mapping probabilities. In other words, the probability of extra dummy nodes above the minimum to restore equal sizes to the cliques is considered to be negligible. Correspondingly the smaller of the two cliques is padded out with dummy nodes until the cardinalities are equal and the generation of SPMs continues as before (Figure 2.3).

It is also worth noting that if some property of the graph structure is invariant under

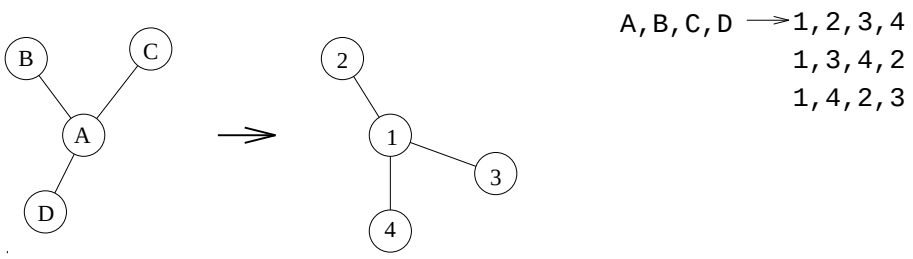


Figure 2.4: Example clique mapping with cyclic order preserved

the possible set of transformations between the two scenes, this information can be used to reduce the size of legitimate SPMs. For example, if the scenes are known to be planar the cyclic order of the external nodes is always preserved (Figure 2.4) between different views (for example in aerial images). In this case we need only explore mappings which also preserve this order. Dummy nodes may still be inserted at any point within the cyclic order of external nodes.

2.6.2 Consistency criterion

Our basic philosophy in constructing a consistency criterion is to use the probability of a labelling as a measure of the quality of match. This approach allows the construction of an objective matching model. To achieve this aim we must construct a probabilistic model of the processes at work in the matching problem.

To construct a consistency measure we begin by computing the probability of each clique matching as specified by the function f. In other words, we are interested in computing the probability of the matched relation Γ_j assigned to the clique $C_j^{(1)}$. As we noted in the previous section, the topologically consistent mappings available for gauging the quality of match are represented by the set of relational mappings from $C_j^{(1)}$ onto G_2, i.e. $\mathcal{P}$. As demanded by the Bayes rule, we compute the probability of the required clique matching by expanding over the basis configurations belonging to $\mathcal{P}$

$$P(\Gamma_j) = \sum_{S_i^k \in \mathcal{P}} P(\Gamma_j|S_i^k).P(S_i^k) \tag{2.8}$$

The models we require are represented in terms of the conditional matching probabilities $P(\Gamma_j|S_i^k)$ and of the joint priors $P(S_i^k)$ for the consistent relations in the set of legitimate

mappings. After Hancock and Kittler (Hancock and Kittler, 1990a) we assume that matching errors exist in the current match f, and that the matching errors on adjacent nodes in the same clique are memoryless. Furthermore we assume that such errors occur with uniform probability distribution. In direct consequence of our assumptions, we may factorise the required probability distribution over the constituents of the relational mapping under consideration. As a result the conditional probabilities may be expressed in terms of a product over label similarity probabilities

$$P(\Gamma_j|S_i^k) = \prod_{k=1}^{|S_i^k|} P(f(u_k)|v_k) \tag{2.9}$$

Our next step is to propose a model of the processes which give rise to erroneous matches. As mentioned above we have assumed that label errors occur with a memoryless uniform probability. If this probability is P_e, then if we have selected the matching SPM S_i^k, the probability that two selected labels will disagree is P_e. Comparison between relations of different sizes is performed by padding the smaller relation with a number of nodes $\mathcal{S}(\Gamma_j, S_i^k) = ||\Gamma_j| - |S_i^k||$ which belong to a no-match category as described in section 2.6.1. These missing nodes, denoted d, correspond to graph corruption and are assigned a probability $P(f(u_k)|d) = P(d|v_k) = P_s$ equal to the probability of node loss through corruption. The confusion probabilities appearing under the product of equation 2.9 are therefore assigned according to the following distribution rule

$$P(f(u_k)|v_k) = \begin{cases} P_s & \text{if } f(u_k) = d \text{ or } v_k = d \\ (1-P_e)(1-P_s) & \text{if } f(u_k) = v_k \\ P_e(1-P_s) & \text{if } f(u_k) \neq v_k \end{cases} \tag{2.10}$$

Combining this distribution with equation 2.9, we obtain an expression for the conditional probability:

$$P(\Gamma_j|S_i^k) = P_s^{\mathcal{S}(\Gamma_j,S_i^k)}(1-P_s)^{\mathcal{C}-\mathcal{S}(\Gamma_j,S_i^k)} P_e^{\mathcal{H}(\Gamma_j,S_i^k)}(1-P_e)^{\mathcal{C}-\mathcal{S}(\Gamma_j,S_i^k)-\mathcal{H}(\Gamma_j,S_i^k)} \tag{2.11}$$

where $\mathcal{C}$ is the size of the larger of the two cliques. As a natural consequence of this distribution rule the joint conditional probability is a function of two physically meaningful

variables. The Hamming distance $\mathcal{H}(\Gamma_j, S_i^k)$ between the assigned matching and the feasible relational mapping S_i^k counts the number of conflicts between the current matching assignment Γ_j residing on the clique $C_j^{(1)}$ and those assignments demanded by the relational mapping S_i^k. The size difference $\mathcal{S}(\Gamma_j, S_i^k)$ counts the number of missing corrupted nodes hypothesised by the mapping. With these ingredients, the resulting expression for the joint conditional probability acquires an exponential character

$$P(\Gamma_j|S_i^k) = K_{C_j} \exp[-k_e \mathcal{H}(\Gamma_j, S_i^k) - k_s \mathcal{S}(\Gamma_j, S_i^k)] \tag{2.12}$$

where $K_{C_j} = [(1 - P_e)(1 - P_s)]^{|C_j|}$. The exponential constants appearing in the above expression are related to the matching-error probability and the corruption probability, i.e. $k_s = \ln\left[\frac{(1-P_s)(1-P_e)}{P_s}\right]$, and $k_e = \ln\left[\frac{(1-P_e)}{P_e}\right]$. The expression may be regarded as providing a natural way of softening the hard relational constraints operating in the model graph. Having developed an exponential expression for the joint conditional matching probabilities, it only remains to specify the distribution of the prior probabilities for consistent relations in the dictionary. Here we adopt a uniform distribution of the available unit probability mass over the set of possibilities $\mathcal{P}$, i.e. $P(S_i \in \mathcal{P}) = \frac{1}{|\mathcal{P}|}$. The final expression for the clique matching probability is therefore (from Eqn. 2.8)

$$P(\Gamma_j) = \frac{K_{C_j}}{|\mathcal{P}|} \sum_{S_i^k \in \mathcal{P}} \exp[-k_e \mathcal{H}(\Gamma_j, S_i^k) - k_s \mathcal{S}(\Gamma_j, S_i^k)] \tag{2.13}$$

Before proceeding, it is important to comment on the structure of the above expression. The most striking and critical feature is that the consistency of match is gauged by a series of exponentials that are compounded over the dictionary of consistently mapped relations. It is this feature that distinguishes it from alternatives reported in the literature (Boyer and Kak, 1988; Herault et al., 1990; Kittler et al., 1993; Li, 1992). Each relational mapping contributes a single exponential to the probability of match. It is this feature that allows our method to operate in a robust manner when the space of relational mappings is large. As we will demonstrate in Chapter 6 compound exponentials of the type defined above offer tangible benefits over linear or quadratic measures in terms of the number of relational mappings accommodated and the label-error probability of the resulting match (Hancock and Kittler, 1993). Moreover, the importance of the different relational constraints is naturally

graded by Hamming distance; relational mappings of large Hamming distance contribute insignificantly while those of small Hamming distance dominate. By gradually reducing P_e, the exponentials appearing in equation 2.13 approach their delta-function limits. This effectively corresponds to subjecting the softened relational constraints operating in the matching problem to a graded hardening. In the limit of vanishingly small error probability the matching probabilities become binary in nature; their role is to effectively count the number of consistently matched relational units. Under these conditions our matching criterion becomes similar in function to the relational distance measure of Shapiro and Haralick (Shapiro and Haralick, 1985). However it is worth noting that in this limit, partially matched relational units do not contribute to the consistency of a match. This is clearly undesirable when the match is poor and few or no fully consistent relations exist. The softening of relational constraints implied in Equation 2.13 alleviates these problems.

We have adopted a very simple distribution rule to specify the label confusion probabilities (Equation 2.10) based purely on a symbolic representation. A number of authors (Boyer and Kak, 1988; Kittler et al., 1993) suggest the use of binary attribute relations to characterize the similarity between label pairs as opposed to Hamming distance. Here we aim to show that the symbolic approach is indeed sufficient to successfully match complex graphs and that it provides advantages in terms of ease of control and a reduction in the number of matching parameters. However our clique matching probability is similar to that of Boyer and Kak (Boyer and Kak, 1988) if we adopt a label similarity based on Gaussian measurement distributions and we assume that the exponentials appearing in Equation 2.13 can be approximated by linear terms (i.e. that the measurement deviations are small).

2.6.3 Defining a Global Criterion

Using our model of the clique matching probabilities, We can define a global criterion of match between any two graphs. There are a number of possible alternatives; we could for example use the joint probability of the clique mappings over the graph. Another alternative is a entropy function of the mapping probabilities (Wong and You, 1985; Boyer and Kak, 1988). The philosophy behind decomposition of the graph into manageable sub-unit is that each sub-unit is representative of the graph as a whole. In effect the cliques are samples of

the complete graph. Following this line of reasoning, the probability of a graph match is best given by the mean probability of the sub-unit matches. This is an approach which is widely employed in relaxation schemes (Hancock and Kittler, 1990a). Our functional based on the match f is given by

$$P(f) = \frac{1}{|V_1|} \sum_{C_j^{(1)} \subset V_1} P(\Gamma_j) \tag{2.14}$$

Accordingly, the MAP criterion we should attempt to maximize is

$$\mathcal{F}_d(f) = \Big\{ \prod_{v_i \in V_1} p(\mathbf{x}_i | f(v_i)) \Big\} \times \frac{1}{|V_1|} \sum_{C_j^{(1)} \subset V_1} P(\Gamma_j) \tag{2.15}$$

We evaluate other forms in the experimental study presented later in Chapter 6.

2.6.4 Applications of the Matching Criterion

The clique matching probabilities and corresponding criterion are influenced by a number of factors, most obviously the current match f and the label error probability P_e. Also of interest are the clique and SPMs. These are dependent on the structure of the graph. These elements can be employed to achieve a number of different optimisation goals.

The simplest use of the criterion in the arena of matching involves the improvement of the matching function f. This application is discussed in Chapter 3. Theoretical issues pertaining to this are discussed in Chapter 5. The criterion may also be used to control the structure of the graph by assessing the impact of changes in the clique and SPMs. This is applied as a method for controlling clutter in the graph in Chapter 4.

2.7 Summary

In this chapter we have developed a method of decomposing a graph into small sub-units and calculating the topologically legitimate set of mappings between these units. With these mappings to hand, probability distributions can be defined for the different possible matching configuration and the probability for a particular match of a unit can be calculated.

Treating these sub-units as samples of the graph as a whole, the global matching probability is defined as the average clique match probability. The MAP estimate of the match probability can then be exploited to achieve a number of graph-matching tasks.

Chapter 3

Discrete Relaxation

3.1 Introduction

In the previous chapter we developed a criterion based on the consideration of structural constraints from the graphs. The criterion is based on probability distributions and is defined over a set of entities. The criterion is defined in terms of a mapping function f which represents the current match between graphs in terms of a many-to-one mapping from G_1 to G_2. The next step in solving the matching problem is then to locate the maximum of the criterion corresponding to the most probable match between the graphs. Discrete optimisation is not a mature field and research is still being conducted into new algorithms (Milun and Sher, 1993) which include simulated annealing and genetic search. Here we will briefly discuss the problems of optimising a discrete function and some of the techniques available to perform the task.

3.1.1 Updating the Discrete Criterion

The main hindrance to the accurate and efficient discrete optimisation of a matching criterion is the function itself. There are two problems which present themselves; these are local maxima in the function and deadlocked updating. The aim of the optimisation phase is to locate the global maximum (which is the largest local maxima) of the criterion, which corresponds to the optimal solution of the matching problem. If there are local maxima present which are smaller than the global maximum this provides a severe test for the

optimisation procedure. Methods which rely on local gradient information will fail in this case precisely because they use local information - there is no information about the large-scale structure of the function (Geman and Geman, 1984). Update methods which wish to overcome the problems of local minima must make use of some global knowledge of the matching criterion. The second problem, that of deadlocked updated is caused by a flat criterion around the current matching configuration. In other words if the function makes use of a coarse measure of consistency, all matching configurations close to the current match could potentially have the same consistency value. In this case the problem is with determining the necessary update to move towards the solution. These difficulties can be overcome either with a global view of the criterion or with a finer measure of consistency

3.1.2 Discrete Optimisation

There are a number of techniques discussed in the literature for optimising a discrete function. The basic approach is the parallel method to gradient ascent(GA) in the optimisation of continuous functions (Hancock and Kittler, 1990a). In this approach we simply choose the label update which results in the maximum increase in the criterion. It is the local nature of the function which determines the direction of label update and the method is susceptible to sub-optimal local maxima. However it is straightforward and efficient if there is only one maximum to find.

Simulated annealing(SA) (Geman and Geman, 1984; Herault et al., 1990) is a far more sophisticated technique, and it is discussed in detail both in Chapter 1 and in section 5.3 in Chapter 5. In essence, the method is able to escape from local minimum traps by allowing updates which both increase and decrease the value of the function. Updates which improve the consistency of the match are always allowed, whereas consistency-decreasing updates are carefully controlled by allowing them with a probability dependent on the change in consistency and the 'temperature' of the system. As the temperature is decreased, consistency-decreasing jumps become more and more unlikely, resulting in an update procedure more and more like the traditional gradient ascent approach. It is this temperature which effectively provides the global information about the state of the labelling (the temperature decreases as the labelling approaches the optimal point) and smoothes out

the sub-optimal local maxima. While this approach provides a technique for escaping local maxima, it is extremely inefficient and converges only slowly.

Genetic search (Fogel, 1994) employs the processes of mutation and selection to generate a population of new solutions to the optimisation problem which are superior in terms of a fitness measure to the initial solution. The mutation process is a random update procedure which is reminiscent of the Metropolis algorithm. Because the randomness of updates, the method is generally very slow to converge but is potentially able to escape from local maxima by virtue of generating consistency-decreasing configurations as part of the population of solutions.

These sophisticated update schemes are necessary when there are problems with local optima and a coarseness in the criterion. However we have developed a fine measure of relational consistency which involves a parameter P_e which has meaning in terms of the current match. This parameter naturally smoothes out local minima of the criterion. Consequently we have adopted gradient ascent as our optimisation method. In this Chapter we demonstrate that the structurally based criterion function defined in the previous chapter when coupled with a simple GA optimisation scheme is able to effectively match graphs under a variety of testing conditions.

3.2 Relaxation

With the average consistency criterion to hand we can iteratively update the mapping function $f : V_1 \rightarrow V_2$ on a node-by-node basis to locate an optimal match. The updating process is therefore effected by replacing one of the node mappings belonging to f by the match that results in the greatest improvement in the value of the MAP estimate of the labelling. This optimisation strategy has the dual advantages of accommodating the persistence of observational information and being realisable by simple gradient ascent. The aim of this decision scheme is to locate the matching configuration that has maximum *a posteriori* probability (MAP) with respect to the available observations. The result used here is derived by Hancock and Kittler in (Hancock and Kittler, 1990a). According to our philosophy, structural information is modelled by the joint prior $P(f)$. Observational

evidence for matching affinity between data node $u^{(1)} \in V_1$ and model node $v^{(2)} \in V_2$ is captured by the single probability of the relevant unary measurement information, i.e. $P(f(u_k)|\mathbf{x}_k)$. The initial configuration of the relaxation scheme is seeded on the basis of the maximum value of $P(f(u_k)|\mathbf{x}_k)$. Updated matches are selected to optimise the following quantity which is proportional to the *a posteriori* probability of the global matching configuration (see Chapter 2, section 2.6.3).

$$\mathcal{F}_d(f) = \Big\{ \prod_{v_i \in V_1} p(x_i|f(v_i)) \Big\} \times \frac{1}{|V_1|} \sum_{C_j^{(1)} \subset V_1} P(\Gamma_j) \tag{3.1}$$

Consider the updating of the label on node $u^{(1)}$. The appropriate GA update rule for this node is

$$f(u) = \arg \max_{v(2) \in V_2} \Big\{ \mathcal{F}(f(u^{(1)}) = v^{(2)}, f(w^{(1)}), \forall w^{(1)} \in V_1, w \neq u) \Big\} \tag{3.2}$$

Since only the label on this node changes and the unary measurements are independent, we need only consider the change in conditional measurement probabilities at the node u itself; i.e.

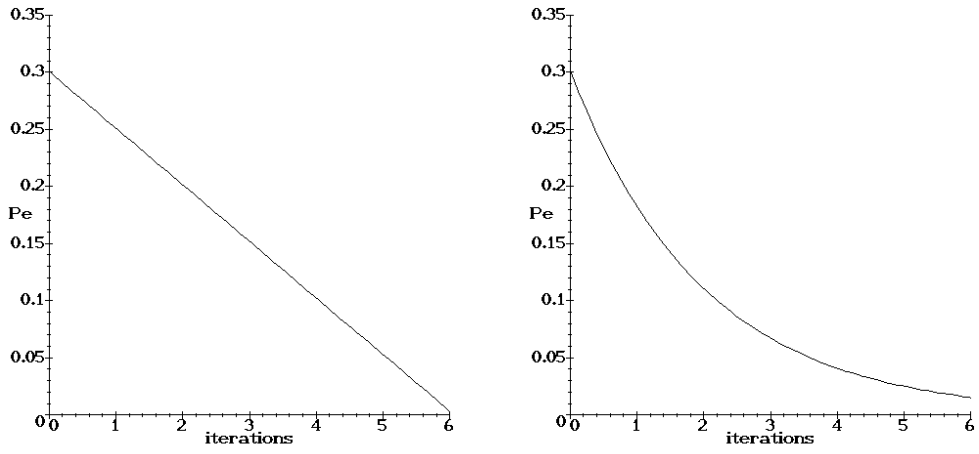
$$f(u) = \arg \max_{v(2) \in V_2} \Big\{ P(\mathbf{x}_u|f(u^{(1)} = v^{(2)}) P(f(u^{(1)}) = v^{(2)}, f(w^{(1)}), \forall w^{(1)} \in V_1, w \neq u) \Big\} \tag{3.3}$$

We now turn our attention to the joint prior $P(f)$. We do not need to evaluate the entire probability, rather we can confine our attention to the portion of the function that is modified by a change in the match of node $u^{(1)}$. Reference to equation 2.14 shows that we consider the contribution of only the cliques which contain $u^{(1)}$. The final update rule is then given by

$$f(u) = \arg \max_{v(2) \in V_2} \Big\{ P(f(u^{(1)} = v^{(2)}|\mathbf{x}_u) \sum_{w^{(1)} \in C_u} P(\Gamma_w) \Big\} \tag{3.4}$$

In deciding on the label on $u^{(1)}$ structural information is drawn from all the cliques surrounding u, which exploits labelling information from up to two graph edges away from the original node.

In the deterministic sequential implementation of the scheme each node is visited for update once per iteration and all possible matching nodes are tried at that location. A number of iterations are required to spread context across the graph.



Linear Exponential Decay

Figure 3.1: Plots of reduction schemes for the label-error probability

3.3 Parameter Control

By adopting a symbolic paradigm in the matching process, we have arrived at a scheme which is economical in terms of the parameters required. Infact, once the unary measurements have been incorporated into initial matching probabilities, the relaxation scheme has just two parameters, both of which represent physically meaningful quantities. The probability of relational corruption P_s is equal to the probability of nodes being lost or spurious nodes being introduced into the cliques and therefore is related to the amount of graph corruption. This quantity reflects a property of the graphs themselves and remains static throughout the relaxation process. The label error probability P_e on the other hand is effectively a control parameter of the relaxation scheme analogous to the temperature of an annealing scheme, but with an interpretation in terms of the quality of the labelling. This parameter should reflect the prevailing level of labelling errors currently in the match (Hancock and Kittler, 1990a). Since the relaxation scheme should iteratively improve the labelling, one strategy is to set it to a initial high value to reflect a poor labelling, and reduce it according to a deterministic schedule to some small terminal value. This corresponds to a graded hardening of initially soft constraints which has the effect of driving out label errors.

Since the labelling fidelity improves with time, the need for a scheme which reduces the value of P_e with each iteration is anticipated. A number of schedules for the reduction of P_e

have been experimentally tested (Plots of the schemes under study are shown in Figure 3.1; they show the value of P_e as a function of iteration number). The first scheme represents a linear reduction in the label error probability with each iteration to a terminal value of zero. In this scheme we anticipate a constant rate of labelling improvement with iteration down to zero final matching errors. The second scheme is an exponential decay of the probability given by the equation $P_e = P_e^{(0)} \exp[-ki]$ where i is the iteration number and k is some empirically chosen decay constant. This scheme is more realistic in terms of the rate of matching improvement. It represents a rapid early increase in matching fidelity which tails off towards zero matching errors, ending at a small terminal value. Finally we have also tested a non-reduction scheme in which P_e is held static throughout the matching process at the same terminal value as that of the exponential reduction scheme. This experiment is performed in order to test the validity of our assumption that reduction is required to drive out label errors. In order to test the effectiveness of these different schemes, matching has been attempted on the different sets of data described in Appendix A. Details of how the matching experiments are performed and the datasets are explained later in this chapter. Here the synthetic data under match contains 60 nodes and has been subject to 50% corruption. Figure 3.2 shows the relative performance of the three schemes in terms of the fraction of correct nodes (the number of correct matches divided by the maximum possible correct matches).

This analysis shows some interesting results. Firstly the exponential scheme is the best in terms of its labelling performance. However there is very little difference between the exponential and the constant schemes. This suggests that the reduction itself is not a key element in the control of P_e, although the reduction scheme is slightly superior to using a constant value of P_e. Further investigation shows that the constant P_e scheme does not perform well when a continuously high value of P_e is used. In comparison the linear scheme performs poorly. This is attributable to the fact that this scheme terminates on $P_e = 0$; when the data is corrupt the state of zero matching errors is not achievable and theoretical investigation (Hancock and Kittler, 1993) has shown the reduction of P_e to zero to be undesirable. Infact investigation shows that the linear scheme performs similarly to it's exponential counterpart until the final iteration with $P_e = 0$.

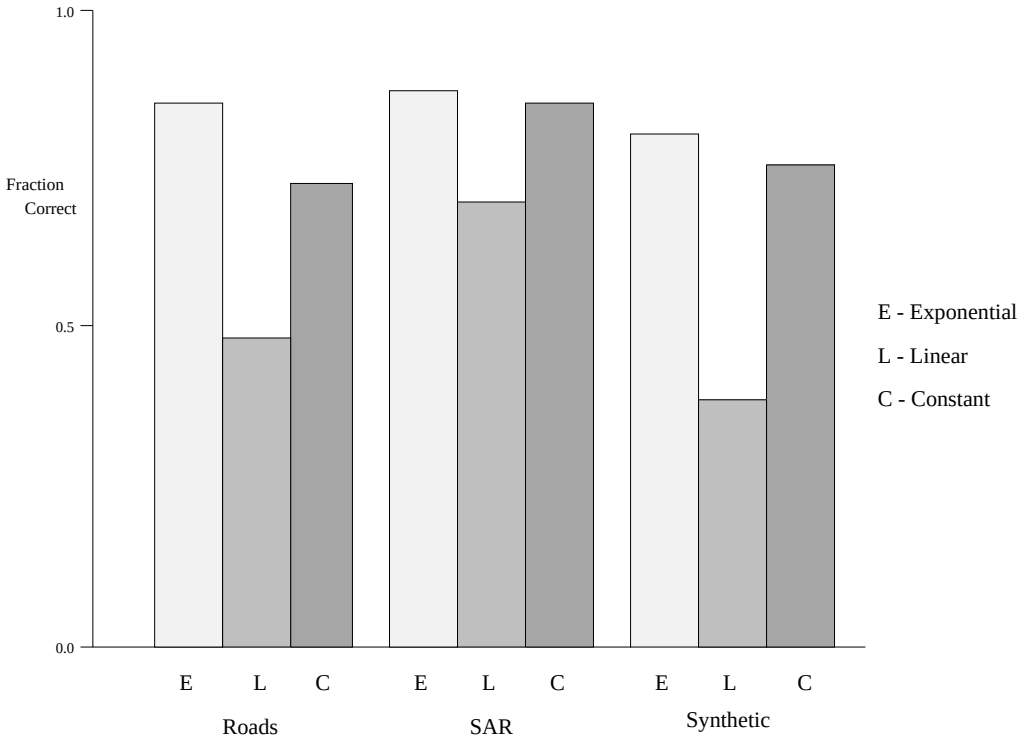


Figure 3.2: Performance of different label-error probability reduction schemes

In conclusion, the key elements to the control of the label error probability are that the schedule terminates on a small but non-zero value. Schedules in which P_e is reduced or remains static give very similar performances, although the results show a marginal advantage to the reduction schemes over the static scheme. We have chosen to use the exponential reduction schedule for the dual reasons that the scheme gives the best performance and we anticipate a faster improvement in labelling in the early iterations when there is more potential for labelling improvement.

The other aspect of the reduction schedule concerns the initial value of the label-error probability, $P_e^{(0)}$. This should reflect the number of initial errors in the labelling, but study of the expression for the constant of the exponential, i.e. $k_e = \ln \frac{(1-P_e)}{P_e}$, forces the condition that $P_e < 0.5$ for k_e to be a positive constant. Above this value, an *increase* in the number of label errors results in an increase in the function $\mathcal{F}_d$. As a result, with $P_e > 0.5$ the discrete update procedure produces the maximum number of incorrect labels. We must confine ourselves to the regime $P_e < 0.5$ and in practice we cannot set P_e according to the demands of the data. The danger of this is that the criterion will be insufficiently smoothed by an artificially

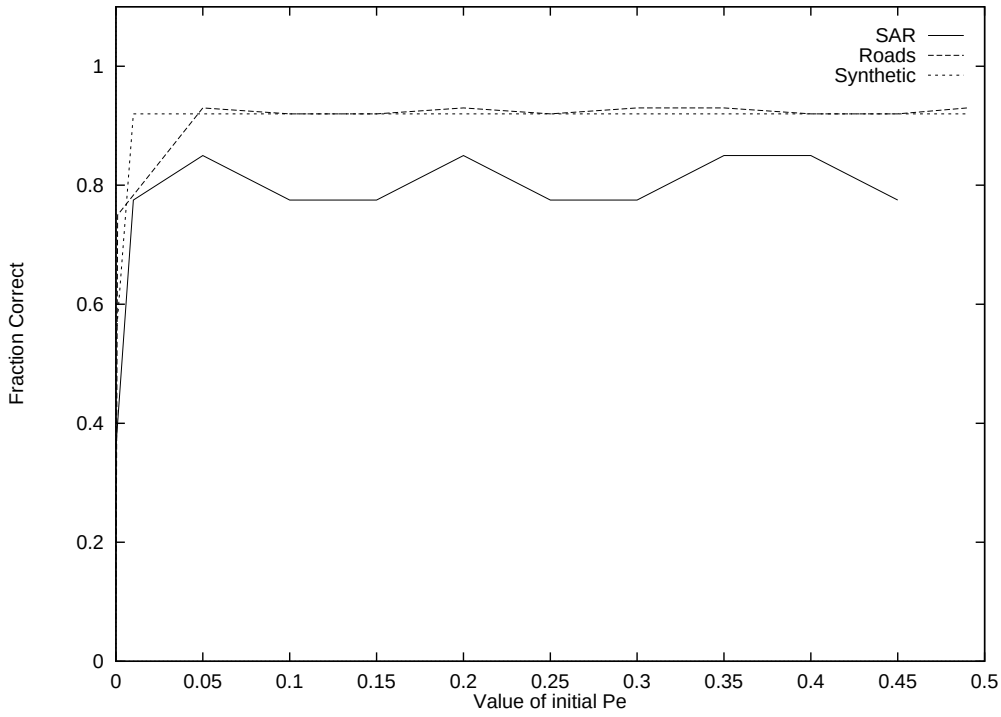


Figure 3.3: Performance for different initial label-error probabilities

low error probability. To investigate this factor we again perform a number of matching experiments, this time over a range of initial values for P_e. Figure 3.3 demonstrates how the matching performance varies with $P_e^{(0)}$ for the exponential scheme.

Study of Figure 3.3 suggests some instability in the final labelling for both the SAR and aerial infra-red data-sets, but it should be noted that this variation corresponds to a change of one correct/incorrect match and is due to the sequential nature of the update process in our implementation. Apart from this there is very little variation over the whole range of values for $P_e^{(0)}$. However the quality of match falls off rapidly if $P_e^{(0)}$ is set initially to zero. This clearly emphasises the need to soften constraints.

This empirical approach, while experimentally based, is effectively a non-rigourous solution to the problem of algorithm control. Another approach is to attempt to extract the optimal value of the label error probability using the information provided by the topology of the graph. More details of this approach are provided in Chapter 5.

3.4 Experimental Investigations

In this section we are interested in testing the effectiveness of our discrete relaxation scheme for matching graphs relevant to computer vision problems.

Two critical elements which affect the performance of all optimisational graph matching algorithms are the level of graph corruption and poor initial matching conditions. These elements reflect the varying quality of data from a scene; uncertainty in the scene can lead to poor measurement information, missing objects and an inability to reliably extract a stable relational representation. With the experiments in this section we intend to make a detailed study of performance under varying conditions of scene corruption and initial labelling quality.

Also of interest with regard to our iterative discrete relaxation algorithm are studies of both the rate of convergence and performance under scene occlusion. As far as the rate of convergence is concerned, this is critical in determining the feasibility of the proposed scheme. Because of the use of large contextual units in the algorithm we anticipate a relatively fast rate of convergence.

Occlusion effects are of interest because of their fundamental importance to the interpretation of three-dimensional scenes from images, where occlusion is a significant effect in the variation of relational graphs between different viewpoints. We anticipate considerable success in this area because the graph is represented in terms of small sub-units, and therefore we would expect the matching to be invariant to occlusion effects provided sufficient structural units remain to provide the required context.

In summary we will investigate the following factors affecting graph matching feasibility in the context of the discrete relaxation matching method over a number of data-sets:

- Random corruption of graph nodes
- Varying quality of measurement information (initial match quality)
- Corruption by occluding portions of the graph
- Rate of convergence

3.4.1 Investigating Scene Corruption

The greatest potential problem facing any matching scheme is corruption of the data under match. Data obtained from a scene is invariably corrupt due to noise and segmentation error. Such uncertainty hinders the matching process and eventually renders it inoperable.

In the context of the data we present here, all the scenes under match are corrupt. In the case of the real-world data the level of uncertainty is not under our control; rather it is a function of the natures of the scenes themselves. For this reason and in order to explore the entire range of data corruption, the bulk of the results are based on synthetically generated data in which the level of error is under experimental control.

3.4.2 Investigating the Effect of the Initial Match

The seeding of the initial match is clearly important to the relaxation scheme as it forms the starting point for subsequent label updates. The labelling is based on a set of unary measurements on the objects under match. It is the effect of the quality of these measurements which is under study here. Poor measurement information leads to initialisation errors from which the relaxation scheme must recover. Variation in the quality of the initial labelling is easily effected in the synthetic data by perturbing the measurement information. The image-based data has a natural degree of variability in the measurements.

3.4.3 Investigating the Effect of Occlusion

Occlusion is generally a feature of the study of 3D scenes but although none of the data here is within this domain, the matching method here is equally applicable to such scenes. For this reason, occlusion is simulated by progressively masking out portions of the image under match.

3.4.4 Investigating the Rate of Convergence

Relaxation schemes vary greatly in the number of iterations required to reach a consistent match. At one end of the scale the stochastic relaxation of Geman and Geman (Geman and Geman, 1984) requires many thousands of iterations. Schemes based on the original

Rosenfeld, Hummel and Zucker probabilistic relaxation scheme (Rosenfeld et al., 1976) rely on spreading weak context across the scene and typically require tens to hundreds of iterations (Price, 1985). Here we investigate the quality of the final match with regard to the number of iterations used.

3.4.5 Experimental Data Types

This subsection describes the various data-sets used in the experimental evaluation of the discrete relaxation technique. For a more detailed discussion of the extraction from images and the formation of graph-structures, the reader should refer to Appendix A.

Road Networks

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 road structure. This data set represents fairly straightforward conditions of low levels of corruption of about 30% of nodes, with relational units of orders 4 at T-junctions and 2 at line-endings. Our matching of the two scenes is based on finding correspondences between the T-junctions and line-endings which delineate the road network. At this point it is worth mentioning that segmentation error will have different effects on the two node-types. Pairs of line-endings are created by line fragmentation, line-ending triples result from junction occlusion and single line-endings are produced by poor junction reconstruction. Spurious T-junctions are less likely since they are produced only as a result of over enthusiastic gap filling. As a result the number of spurious line-endings is much greater than the number of spurious T-Junctions. For this reason we can anticipate very different qualities of matching in the two cases, and hence we will differentiate between them in presenting our experimental results.

The initial matches have been seeded using the probability model described in Appendix A which attempts to capture some of the systematics of the line segmentation process. The initial matches are determined using the angle and length information of the roads which form a junction.

Three separate data sets are available; they comprise of the road network viewed from different altitudes together with a digital map of the road network. The original images are

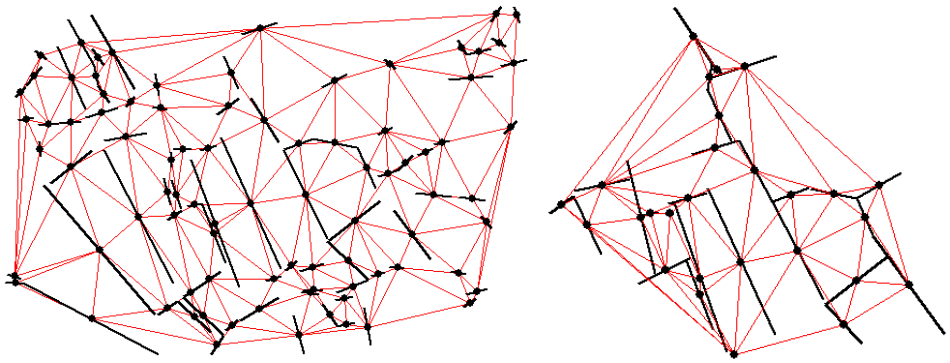


Figure 3.4: Graph structures generated by SAR data and map

presented in Appendix A and the graphs under match are topologically equivalent to the road networks themselves. The map provides uncorrupted ground-truth information.

SAR Matching

Again Appendix A describes the extraction of suitable graph structures for matching. Figure 3.4 shows example graph structures generated by the Voronoi tesselations for the map model and SAR data respectively. There are several features of these graphs that merit special mention. In the first instance, there is considerable variation in the sizes of the clique for the different nodes. The smallest contains only 4 nodes while the largest contains 10 nodes; this means that the order of the largest symbolic relation exploited in the optimisation phase of the discrete relaxation is 10. It is also clear that the data graph suffers from both relational drop-out and relational contamination; there are both spurious and missing arcs. In the region covered by the model graph there is a rate of node corruption of some 43%. There are significant topological differences between the two graphs to be matched.

The initial matches between the linear segments extracted from the SAR data and their map representation are established on the basis of the angular affinity as explained in Appendix A. While there may be a considerable quantity of information remaining untapped by this model, accuracy at this stage is not our primary concern. We wish to demonstrate that the relaxation scheme can recover from poor initial configurations.

The experimental matching study is based on 95 linear segments in the SAR data and 34 segments contained in the map. However only 23 of the SAR segments have feasible

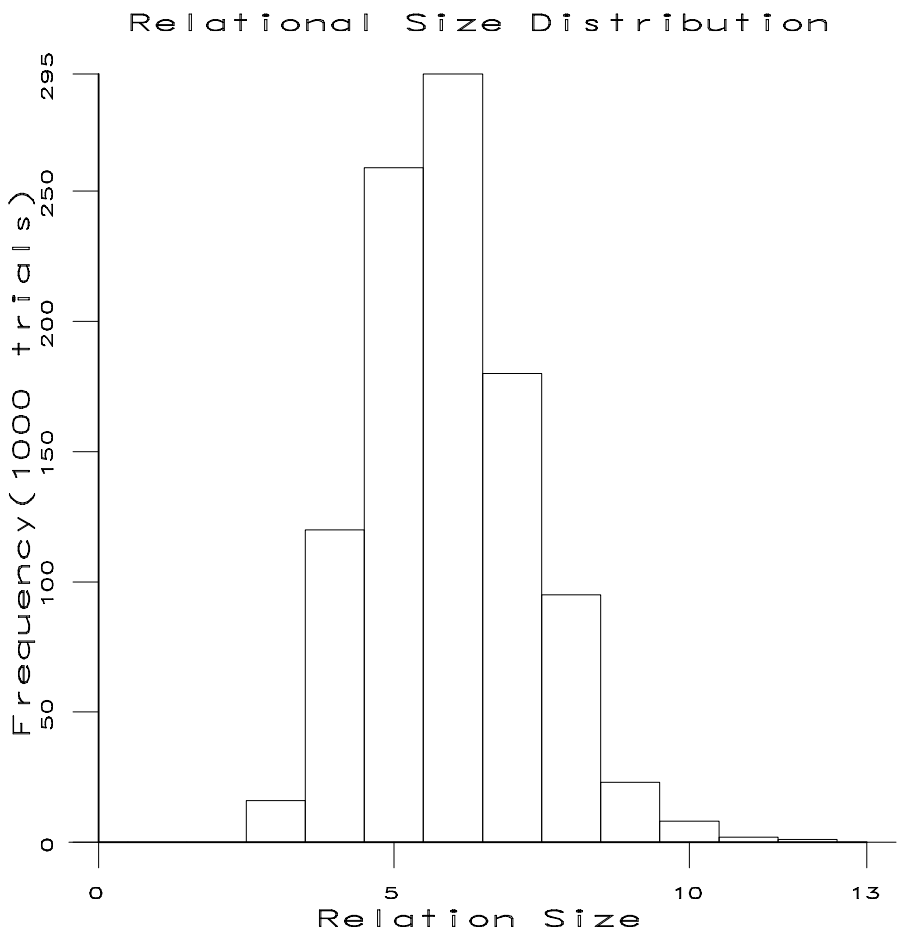


Figure 3.5: Distribution of relational cardinality in a Delaunay graph

matches within the map representation.

Synthetic Data

Because of the limited quantities of available data, while the experiments above provide a challenging real-world application, they do not represent an ideal vehicle for demonstrating the performance characteristics of this method. It is important to provide a study of the performance of the matching algorithm under controlled and varying levels of corruption and measurement uncertainty. We wish to explore the domain over which the scheme is effectively operable, both in terms of corruption of the graph topology and poor initialisation.

In order to embark on this study, graphs have been generated consisting of a number of lines with random positions and orientations. The patterns have been given a relational abstraction by seeding a Voronoi tessellation from the line centre-points and computing the

associated Delaunay graph. The nodes of the graph are therefore the random points, while the arcs indicate that the associated Voronoi regions are adjacent to one-another. Figure 3.5 shows a histogram of the number of Delaunay neighbours for each dot in the graph; it is these neighbourhoods that form the relational units or cliques in our matching experiments. The mode of the histogram occurs at 6 Delaunay neighbours, however, there are clearly a number of relations of cardinality as high as 9 and 10. By randomly adding and subtracting lines from the patterns we can simulate the effects of noise, clutter, segmental dropout in the matching process. In addition by adding Gaussian noise to the line angles, we can simulate poor initial measurements which allow us to control the quality of the initial labelling.

3.5 Performance and Sensitivity Analysis

In the following results, the matching scheme is evaluated according to a simple performance measure:

$$F_c = \frac{N_c}{N}$$

where N_c is the number of correct matches which are found and N is the possible number of correct matches available.

As an example, if the data is 50% corrupt, only half of the nodes have feasible matches. If all 50% of these were correctly matched by the algorithm, the performance measure $F_c = 1.0$. On the other hand if 25% of all the nodes were correctly matched then $F_c = 0.5$. Matching performance is therefore measured over the interval $[0, 1]$.

3.5.1 Scene Corruption

As mentioned earlier, our main experimental vehicle here is the synthetically generated data. We are interested in studying the effect of different levels of scene corruption on the matching process. To commence, Figure 3.6 shows the effect of corrupting nodes on the relations used in the matching process.

The deletion of one node affects the composition and structure of a number of adjoining cliques, amplifying the effect of corruption in the Delaunay relations. For example when

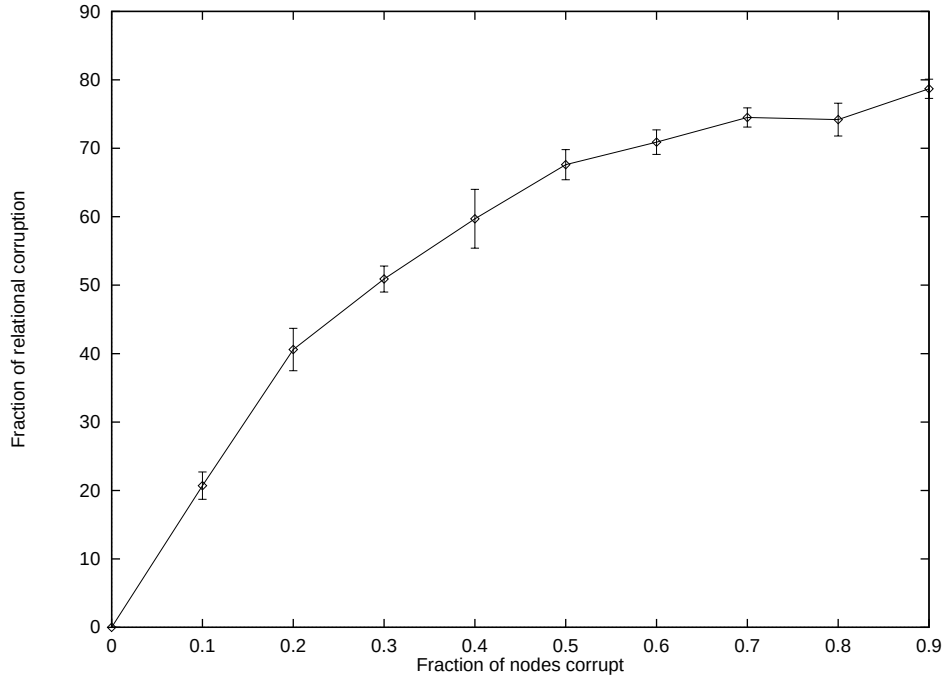


Figure 3.6: Effect of node corruption on relations in a Delaunay graph

30% of the nodes are corrupt, 50% of the average Delaunay relation is altered.

Figure 3.7 shows the fraction of correctly matched nodes as a function of the level of random graph corruption as measured by the fraction of corrupted graph nodes, for the case of synthetic data. The dotted line gives an indication of the initial matching conditions, i.e. the measurement information was sufficient to allow the initial correct matching of approximately 50% of the matchable nodes. Once the level of graph corruption reaches 60% there is little or no improvement to be gained from the application of the relaxation scheme. Provided that the level of graph corruption does not exceed 10% of the nodes, then an almost perfect match is recoverable. The performance rapidly drops off above 50% corruption.

Also marked on the graph are two points corresponding to the road network and SAR data sets. The corruption levels of these data sets are calculated from ground-truth matches by looking at the number of unmatchable elements within the portion of the data graph covered by the model graph. External clutter in the portion of the scene not included in the model graph is ignored for the purposes of this calculation; the significant factor is the level of relational corruption among the elements to be matched. The numbers of these unmatchable entities gives a rough indication of the level of scene corruption. It should be

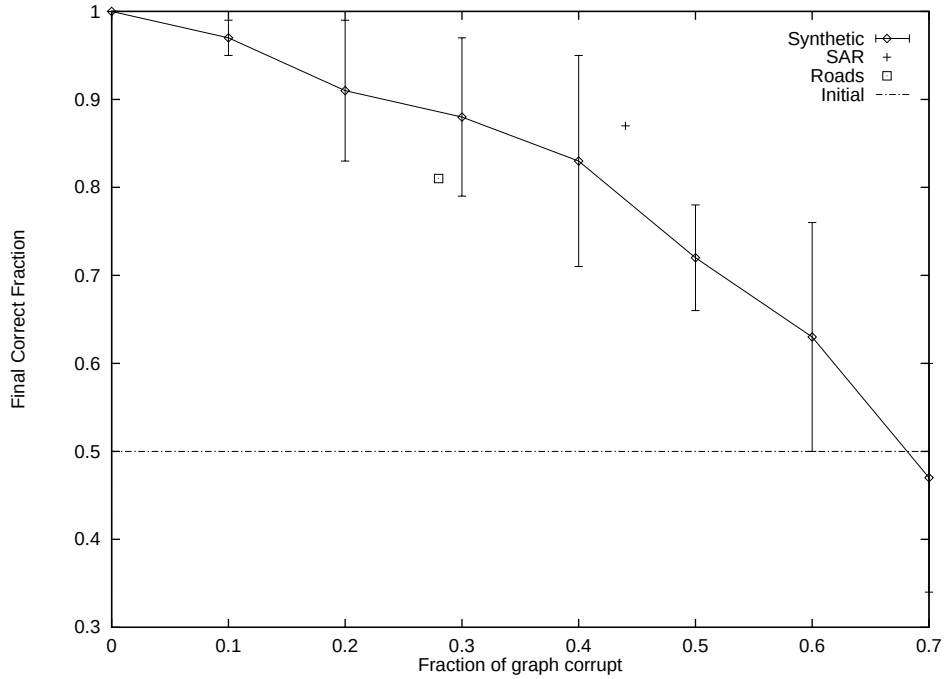


Figure 3.7: Matching under graph corruption using discrete relaxation

stressed however that it is not feasible to compile error-bars for these points.

The road network data has a corruption level of 28% with 49 T-junctions and 6 line-endings initially correctly matched. After application of the discrete relaxation scheme the results improve dramatically. Of the T-junctions 82 of 97 possible correct matches are found. For the final line-ending interpretation, 33 of a possible 45 match correctly. The match is shown graphically in Figure 3.8, the top picture corresponding to the initial matches and the bottom displaying the final matches after relaxation.

The SAR data has a higher level of corruption at 44%. Initially there are 11 correct matches. After application of the method increases the number of correct matches to 20 of a possible 23 matches. The number of matching errors is still substantial but these mainly represent unmatchable scene clutter. A number of techniques for removing these spurious clutter segments are developed in Chapter 4. Figure 3.9 shows pictorially these correct matches; the incorrect matches have been removed for display purposes.

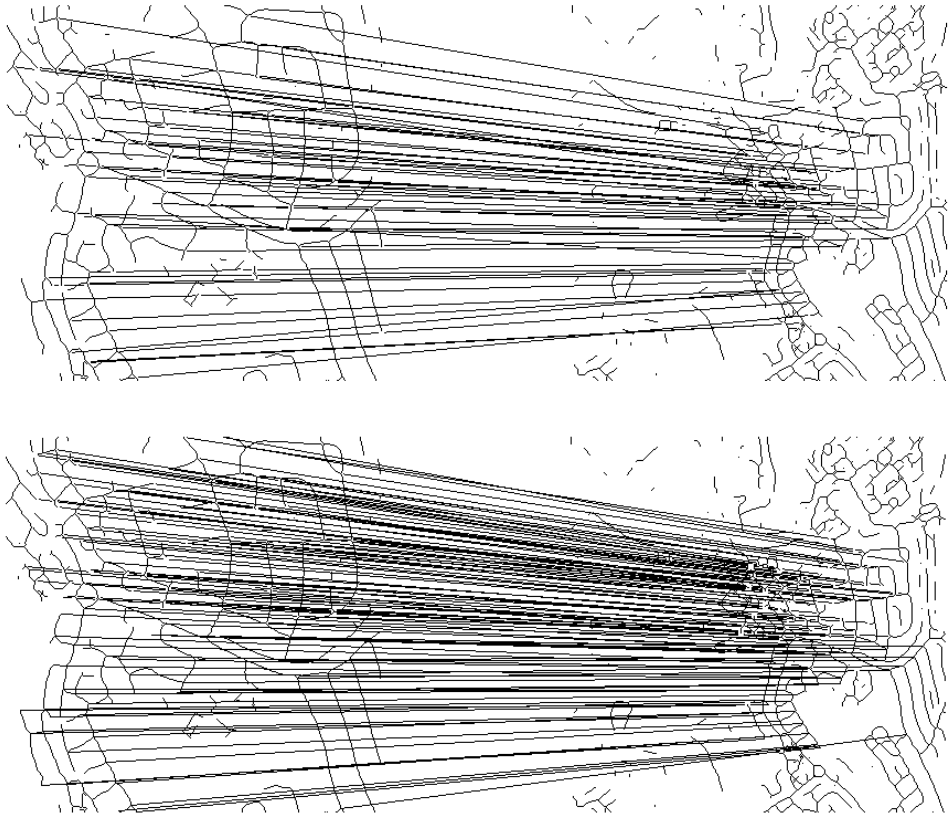


Figure 3.8: Matching of Road Networks: Initial(top) and final(bottom) matches

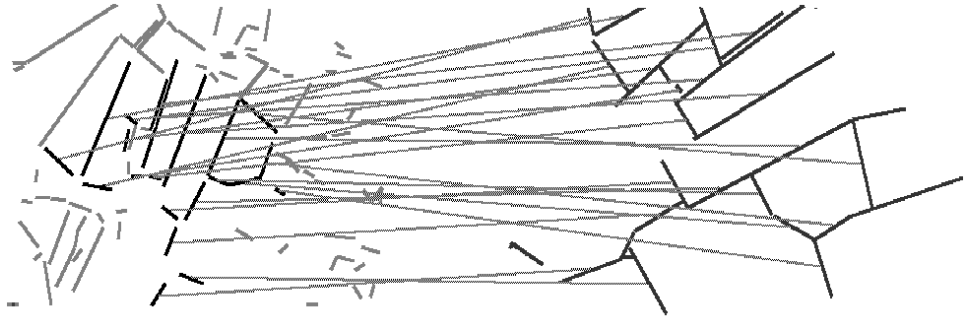


Figure 3.9: Matching of SAR line segments (incorrect matches removed)

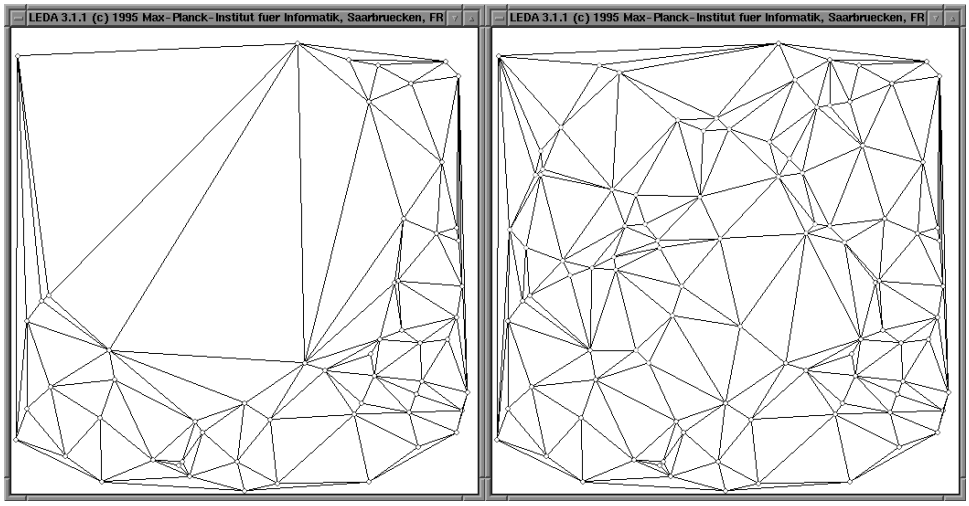


Figure 3.10: An example occluded scene

3.5.2 Occlusion

The results presented here are designed to show the effect of occluding portions of the data graph. The study here is based on 100 synthetic nodes which are then occluded using a circular mask across the image. An example subject of occlusion is shown in Figure 3.10. Initially approximately 50% of the nodes correctly match. Figure 3.11 shows the result of relaxation on data with varying levels of occlusion.

This figure demonstrates that there is little effect for levels of occlusion up to 80%. Subgraph matching proceeds as effectively as the full match. This is an exciting feature of the graph decomposition approach; the results here suggest that it is feasible to match scenes with up to 80% of the structure occluded by a foreground object. Above this level the matching algorithm rapidly becomes completely ineffective.

3.5.3 Initial match

Figure 3.12 shows the labelling improvement over a number of iterations for an initially poorly labelled synthetic matching problem. In the problems under study here the graphs under match are of identical topology and only the initial match is corrupt. This figure clearly shows the ability of the scheme to recover from a very poor labelling state. Figure 3.13 expands on this observation by analysing the quality of the final match over a range of initial labellings. This figure illustrates that provided there are no other sources of error

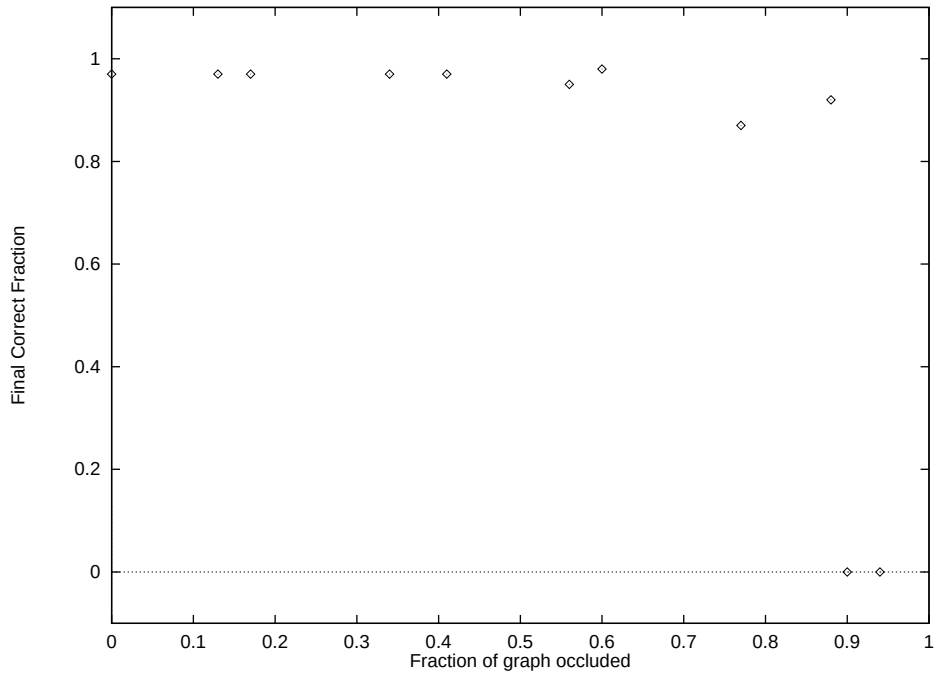


Figure 3.11: Matching under occlusion using discrete relaxation

present, then a fully consistent global match can be recovered if as few as 10% of the initial matches are correct, and even when only one match is initially correct, as much as 80% of the correct match is recovered.

3.5.4 Rate of Convergence

Here we wish to investigate the minimum number of iterations required for our discrete relaxation scheme to converge to a stable and accurate match. In order to do this the algorithm has been run with varying numbers of iterations, while the reduction scheme for P_e is modified each time. This modification is such that it allows the initial and terminal values of P_e to remain the same while the number of iterations determines the number of points sampled in between. For example, if three iterations are used, both the predetermined initial and final values are used as well as one point sampled in between. The reduction proceeds according to the exponential decay law in section 3.3. Figure 3.14 shows the effect of varying numbers of iterations on the matching performance.

It is clear that there is some variation in the number of iterations to convergence between the different datasets. For the SAR data, three iterations are sufficient to locate the optimal

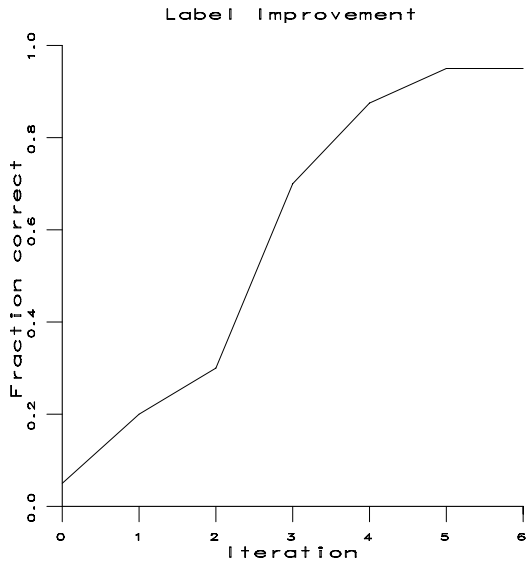


Figure 3.12: An example of the iterative improvement of labelling

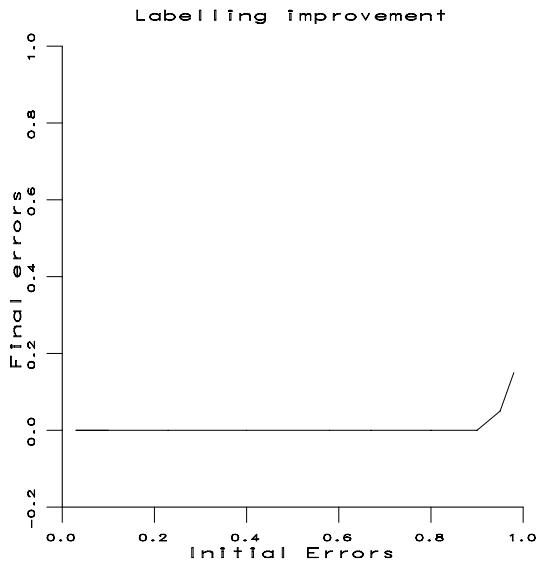


Figure 3.13: Matching under varying initial match corruption using discrete relaxation

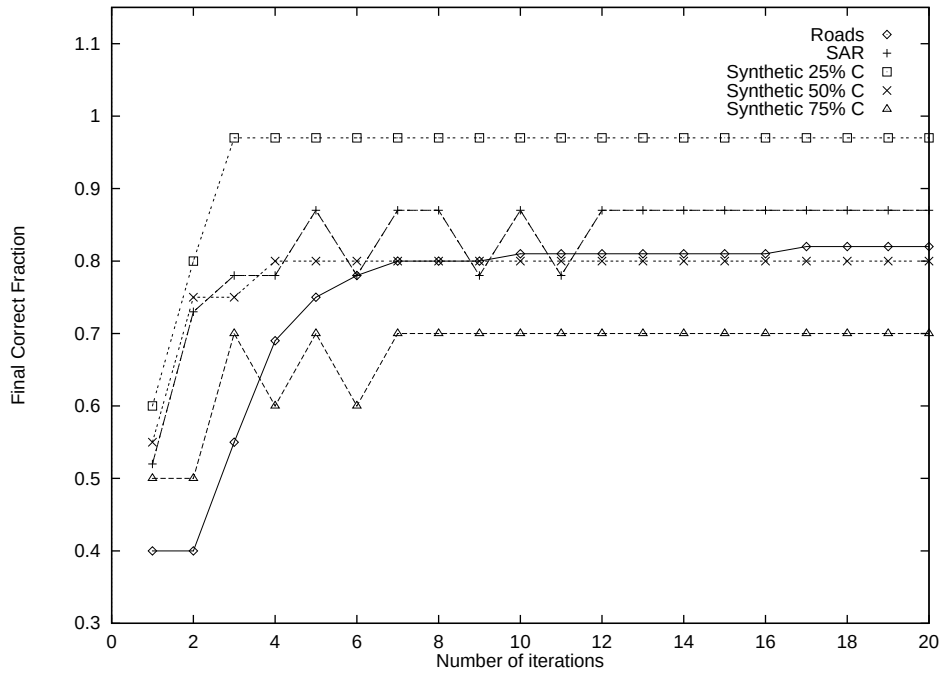


Figure 3.14: The effect of varying the number of iterations

match, whereas the road network data requires 17 iterations to find the best match. This seems to be attributable to the much larger number of matchable nodes the road data; more iterations are required to spread the required contextual information across the network before optimal performance is achieved. However by far the majority of the labelling improvement occurs in the first 7 iterations.

3.5.5 Comparison of MAP and Configurational Relaxation

In our MAP matching philosophy, measurement information persists throughout the matching phase in the form of conditional measurement densities. Much recent interest in the field has been focussed on techniques which bring in measurement information at all stages of the matching process. It is interesting to speculate how great a rôle the measurement densities play in our algorithm, and to what extend symbols alone can be made to enforce a consistent match. To this end we have studied the comparative performance of the MAP scheme and a configurational relaxation scheme which optimises the joint probability $P(f)$ only (Figure 3.15; points labelled 'MAP' and 'Config.' respectively). In this case the function we are interested in is given by

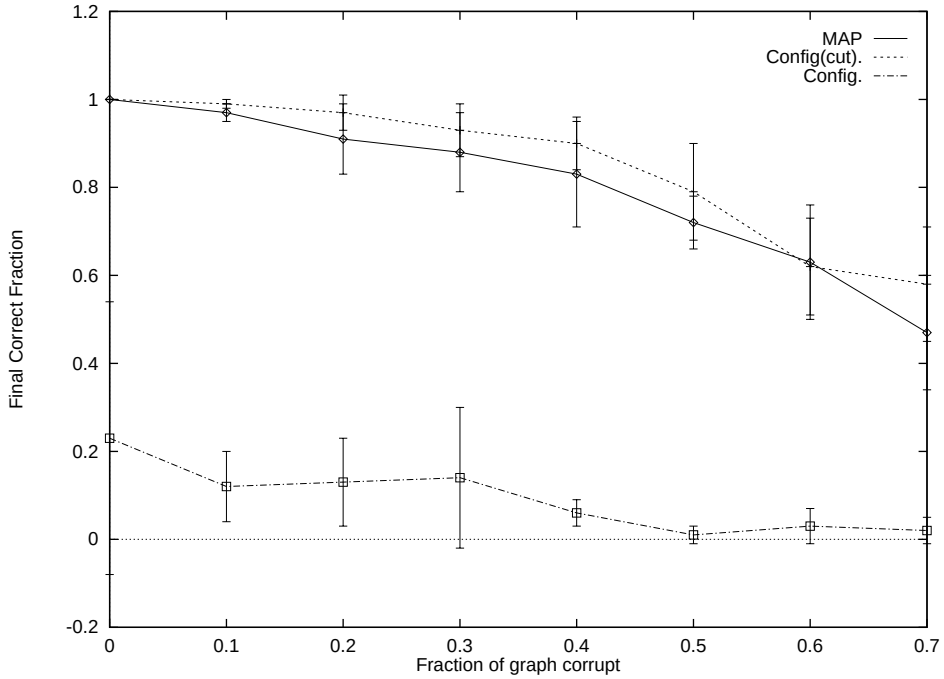


Figure 3.15: Comparison between MAP and Configurational relaxation

$$\mathcal{F}(f) = \frac{1}{|V_1|} \sum_{C_j^{(1)} \subset V_1} P(\Gamma_j) \tag{3.5}$$

The study is performed with synthetic data over a range of levels of corruption (see Section 3.5.1).

Examination of the performance of the configurational relaxation scheme reveals that the matching is completely ineffective at all levels of corruption. The conclusion is that it is a combination of the constraints on the match provided by the unary measurement information and the structural information that allows effective matching to take place, since neither alone are able to locate the correct match.

One possible explanation of the poor performance of the configurational scheme is that the correct matches are swamped by relational noise from background configurations (In Hopfield associative memories it is this effect which limits the storage capacity). One way to investigate this effect is to use the unary measurement information to disallow node matches with very low initial probabilities. By applying this rule we can drastically reduce the space of mappings which need to be explored. A configurational scheme with a pruned configuration space is also displayed on Figure 3.15 (the "Config(cut)" plot). Here a match

$f(u) = v$ has been disallowed if $P(f(u) = v|\mathbf{x}) < 0.01$. From the figure it can be seen that this configurational scheme performs identically to the MAP scheme within the limitation of experimental errors. This provides some confirmation that it is the size of the space of legitimate mappings which limits the performance of a structural approach.

3.6 Conclusions

These simulation experiments demonstrate the effectiveness of the described relaxation scheme under conditions of poor initialisation which is the assumption under which our criterion was developed. Figure 3.13 indicated that a perfect labelling can be recovered when only 10% of the nodes are correctly labelled, demonstrating the powerful use this scheme makes of relational information, and experimentally justifies our use of an explicit model of the role of errors in the matching process.

An encouraging feature of these results is a tolerance to moderate levels of structural corruption. The quality of the final labelling is not significantly affected by levels of corruption up to 20%. However at 30% corruption the performance begins to degrade, until at 60% corruption and above little improvement in the initial labelling is seen. Reference to Figure 3.6 shows that at this level, relational corruption is 70% and over. At this level the constraints which facilitate the matching process are so corrupt that they no longer provide useful information.

In summary the main conclusions of this analysis are as follows:

- We have show that the MAP discrete relaxation scheme developed in Chapter 2 is extremely effective at recovering from poor initial matches due to corruption of unary node measurements. The scheme can recover from 90% mis-labelling to locate a fully correct match. Even at 98% initialisation errors, 80% of the correct match is still recovered.

- The algorithm has also been shown to be tolerant to corruption of the scene graph due to missing or spurious nodes. A 90% correct match can be recovered with 20% corruption. The performance degrades as the structural corruption increases, and at 60% to 70% corruption there is little to be gained by the application of the method.

- We have also shown the method to have a large tolerance to occlusion of portions of the data graph; a fully correct match is located at up to 85% occlusion of the original graph. Above this level the matching is completely inoperable.

- Empirical studies of the rate of convergence to a solution suggest that typically the algorithm is close to the solution after 7 iterations. However, some matches are not located until the 17th iteration for one of our data-sets.

Finally, scene clutter has not been identified by the algorithms presented in this chapter. Incorrect matches still persist due to elements in the data graph which have no feasible match. The next chapter is concerned with labelling and removing such clutter.

Chapter 4

Controlling noise and clutter

4.1 Introduction

Clutter elements are the inevitable result of the imaging and segmentation process on realistic scenes in pattern matching tasks. Scene elements may be missing or fragmented and there may be significant numbers of extraneous objects. In severe cases, extraneous elements may permeate bona-fide scene structure. Indeed it was the presence of these noise objects which motivated us to adopt an inexact approach to the graph matching problem in Chapter 2. These objects inevitably hamper the matching, rendering it susceptible to error. Part of the matching process then must be to identify spurious elements in the scene under match, label them as such and isolate their effect on the matching of the uncorrupted scene elements.

The discrete relaxtion matching technique developed in the last chapter is capable of locating the best match between two graphs, but it has no capacity to either identify or remove nodes corresponding to scene noise. In this chapter we will examine a number of different methods to implement this capability.

As before we are interested in the case when there is insufficient unary measurement information on the elements alone to identify noise. We must therefore turn to contextual information to provide the necessary clues. Examination of the literature reveals three separate approaches to the problem of labelling noise using contextual information. Common to the three approaches is the concept of using relational consistency to drive the filtering out of corruption. The simplest approach is the constraint filtering technique which iden-

tifies inconsistent matches and rejects them as noise. This was the philosophy behind the association graph of Barrow and Popplestone and maximal clique searching. The idea also underpins Waltz's (Waltz, 1975) discrete constraint filtering. Basic to the constraint filtering technique is the idea that the true structure of the scene is relationally consistent with the corresponding match, whereas random noise elements have no such structure. Elements can therefore be filtered out based on whether the set of objects currently residing in the graph form a consistently matched group. Inconsistent elements are labelled as corruption. The graph is thus broken up into a number of sets of mutually consistent nodes and a set of excess inconsistent noise elements. This is exactly the idea behind the maximal clique approaches to matching of (Barrow and Burstall, 1976; Horaud and Skordas, 1989; Jones and Wong,).

The second approach falls firmly within the optimisation framework. It involves the addition of another label to the set of possible matches. This label represents a null or no-match category to which suspected noise elements can be assigned during the matching process itself. An affinity with the null label is then calculated for nodes in a similar fashion to other labels, so the null label process becomes incorporated in the optimisation phase of the labelling. This approach is developed in (Wilson and Hancock, 1993b; Kittler et al., 1993).

A third possible method involves the use of an active graph. In this approach an element which is thought to be noise is removed from the graph and the relations (and therefore the edges in the graph) are reconfigured. The structural similarity between the new graph and a model graph is then computed and compared with that of the previous graph. An improvement in the similarity without the element leads to the classification of that element as noise. This technique has not previously been addressed in the literature, although methods involving modification of the graph have been studied (Messmer and Bunke, 1994; Tsai and Fu, 1983; Sanfeliu and Fu, 1983). However these methods associate arbitrary costs with modifying the graph rather than using the graph structure itself to gauge the effect of a structural modification.

In this chapter we develop these three different schemes for controlling the labelling of noise in graph matching. These schemes are a constraint filtering approach, an optimisation

method and a novel relational clustering technique with dynamic graphs. A comparative study of the effectiveness on both SAR data and synthetic graphs is also undertaken.

4.2 Constraint Filtering

The classical constraint filtering approach has attractive features; we need neither incorporate the possibility of null-matched elements into the label probability distributions, nor do we require prior knowledge of the number of unmatchable elements expected to be present in the data. This tactic limits the number of matching parameters required to a minimum. The softening of relational constraints implied in Equation 2.13 enables the matching process to accommodate these erroneously matched segments while still locating the most consistent match.

If the graphs under match are uncorrupted, we would anticipate a final match which is completely consistent over the whole graph. We have already demonstrated in Chapter 3 that the discrete relaxation matching process is capable of achieving this when graph corruption is not present. In a realistic case in which corruption is a significant factor, this potential area of consistency is broken up into smaller patches, the size of which is limited by the possibility of constraint corruption. Unmatchable elements, on the other hand, have no consistent interpretation and are therefore unlikely to form patches of consistency.

We commence by forming a new graph $G'_1 = (V'_1, E'_1, R'_1)$ which contains the consistently labelled portions of G_1. To form G'_1 we first eliminate arcs whose mapping does not appear in G_2

$$E_1{}' = \{(u_1, u_2)|(f(u_1), f(u_2)) \in E_2{}'\} \tag{4.1}$$

The discarded arcs from G_1 have no consistent interpretation in G_2 and are therefore considered to be the artifact of graph noise and corruption. We then remove disjoint nodes which are no longer connected by an arc; these nodes have no support from the relations in the graphs and are therefore considered to be unmatchable. The new node-set is given by

$$V_1{}' = \{u|(u, v) \in E_1{}', v \in V_1\} \tag{4.2}$$

and the remainder of the nodes are assigned to the null category, i.e. $\forall u \notin V_1{}' [f(u) = \phi]$. The graph $G_1{}'$ now consists of a number of consistent regions H_i in which the nodes are

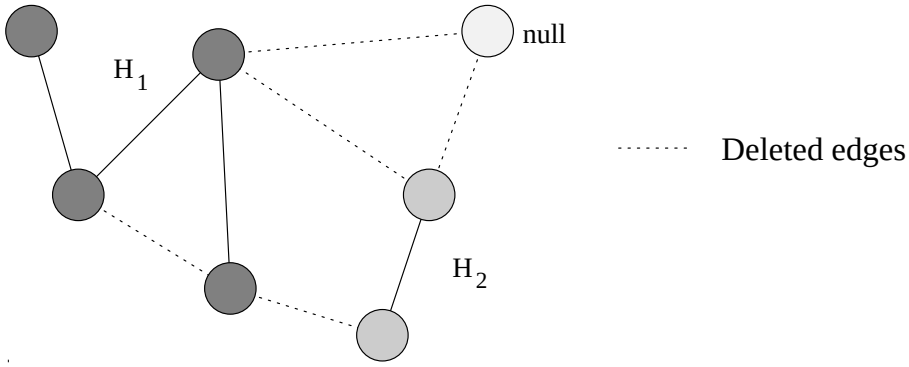


Figure 4.1: Example graph division into regions

connected to each other, but not to any node from a different region. The node-set $V_2{}'$ consists of the union of these regions, i.e. $V_2{}' = H_1 \cup H_2 \cup \ldots \cup H_n$. Formally,

$$u \in H_k \Leftrightarrow v \in H_k \vee (u, v) \in E_1{}'$$

If correct matching were the only process which generates consistency, then $G_1{}'$ would contain only the correctly labelled portion of the graph. However, a small amount of spurious consistency is generated from local isomorphism between erroneously matched segments and regions of G_2. In other words, some edges which correspond to scene clutter can still find consistent matches in G_2 because the local graph structure is similar to a sub-graph of G_2. The probability of this random match occuring reduces as the size of the consistent region increases. We can therefore gauge the consistency of region H_i via a quantity $N_i = |H_i|$ and set a rejection threshold T. If $N_i \leq T$ then all nodes in H_i are assigned to Φ.

In practical matching applications the regions of spurious consistency are small, typically much smaller that correctly labelled regions, and in this case it is sufficient to set the threshold T to remove out these regions without eliminating valid matches. For example in the experiments on corner graphs presented at the end of this Chapter, $T = 3$ is used. However if the level of corruption is very high the consistent regions begin to have sizes similar to the regions of noise and the algorithm becomes inoperable.

4.3 Noise Rejection by Optimisation

In the optimisation approach to clutter identification, we augment the label set of possible matches with an additional null category, i.e. $V' = V \cup \phi$. We must then additionally account for the possibility of null labels occuring in the matched realisation of a clique and suitably modify the distribution of label probabilities specified in equation 2.10. Again we hypothesise a uniform probability of null labels but in contrast to the label-error probability it remains constant throughout the relaxation process, reflecting the anticipated amount of node corruption in the graph. We denote this noise probability with P_ϕ. The new distribution rule is given by

$$P(f(u_k)|v_k) = \begin{cases} P_s & \text{if } f(u_k) = d \text{ or } v_k = d \\ P_\phi(1 - P_s) & \text{if } f(u_k) = v_\phi \\ (1 - P_e)(1 - P_s)(1 - P_\phi) & \text{if } f(u_k) = v_k \\ P_e(1 - P_s)(1 - P_\phi) & \text{if } f(u_k) \neq v_k \end{cases} \tag{4.3}$$

As before, in the case when a label has no match because of the different sizes of cliques under comparison, dummy nodes are assigned with a probability P_s. The corresponding clique probability is

$$P(\Gamma_j|S_i) = K_{C_j} \exp[-k_e \mathcal{H}(\Gamma_j, S_i) - k_s \mathcal{S}(\Gamma_j, S_i^k) - k_\phi \mathcal{N}(\Gamma_j)] \tag{4.4}$$

The exponential constants are now given by

$$k_s = \ln\left[\frac{(1-P_\phi)(1-P_s)(1-P_e)}{P_s}\right]$$

$$k_\phi = \ln\left[\frac{(1-P_e)(1-P_\phi)}{P_\phi}\right]$$

$$k_e = \ln\left[\frac{(1-P_e)}{P_e}\right]$$

while the multiplying constant is given by $K_{C_j} = [(1 - P_e)(1 - P_s)(1 - P_\phi)]^{\mathcal{C}}$.

This new distribution rule implies that the joint conditional probability is now a function of three quantities; the Hamming distance $\mathcal{H}$ and size difference $\mathcal{S}$ as before, and the number of null labels present $\mathcal{N}$. The clique probabilities are therefore naturally graded by the number of null labels present.

It should be stressed that although the no-match probability P_s and the node corruption probability P_ϕ are related, they are not identical. Figure 3.7 in Chapter 3 emphasises this point; the node corruption level is amplified in the level of relational corruption. The relationship between these two quantities is highly dependent on the type of relations employed. The major disadvantage of this approach stems from the need to know apriori the level of corruption present.

The optimisation phase of the matching proceeds as before, apart from the additional evaluation of the probability of a null label at any node. It is important to point out that, while noise elements are labelled as such, the structure of the graph remains unaltered; there is no model of how extraneous or missing elements affect the relational structure in the graph.

4.4 Graph reconfiguration

Different approaches described in the literature incorporate a range of different approaches to relational inexactness. For example the logarithmic conditional information of Boyer and Kak (Boyer and Kak, 1988) captures attribute deviations using Gaussian probability distributions over the measurement space. These distributions are aimed at modelling noise at the level of attributes rather than at the level of symbolic corruption. At the structural level, Shapiro and Haralick (Shapiro and Haralick, 1985) allow the insertion of dummy nodes with no associated penalty, to accommodate the effects of noise and segmentation error. This contrasts with the previous approach in which dummy nodes are incorporated directly into the cost function with an associated probability distribution. The simulated annealing method of Herault et al (Herault et al., 1990) and the graph-edit approach of Messmer and Bunke (Messmer and Bunke, 1994) also associate a cost function with the consistency of match.

Despite the variety of models of relational inexactness, all the techniques described above exploit a static representation of the relationships between objects under match. Spurious elements in the graph contaminate the information supplied by true graph relations, even after the element has be recognised as noise. Under conditions of extreme scene clutter

this can prove to be a severely limiting factor in matching performance. When true scene objects are surrounded by high levels of clutter, little of the original structure is preserved in the data. It then becomes difficult to elicit meaningful relationships from the scene and the matching process only meets with a limited degree of success.

In order to overcome the limitations imposed by high levels of scene noise, we need to additionally attempt to reconstruct the relational structure of the scene when an element is discarded as noise. We must therefore adopt a dynamic representation of the scene objects and their relationships (Messmer and Bunke, 1994; Sanfeliu and Fu, 1983). When noise elements are identified, they are deleted from the scene data and the corresponding graph is recomputed without the object; in this fashion relational structure is restored as the extraneous entities are identified. Eventually the intrinsic relational structure of the model is recovered.

4.4.1 A Dynamic Model of Topology

The task of reconfiguring the graph is one of rejecting noise elements based on the relational structures they form with the rest of the scene graph. As such, it is similar in many ways to the clustering ideas of robust statistics, which are often used to reject a set of outliers to a measurement probability distribution or a uniform coordinate transformation. In this case our aim is to locate entities which form a consistent structural cluster which satisfies the constraints supplied by the model (Henderson, 1990). The MAP criterion developed earlier (Equation 2.15) may now be regarded as a probability distribution which can be used to reject a population of relational outliers. This approach has some features in common with the constrained clustering technique of Rose, Gurewitz and Fox (Rose et al., 1993) but rather than exploiting a Euclidean distance measure, we rely on a Bayesian framework to provide our probability distributions.

Suppose we wish to assign a graph element u from G_1 to the set of outliers Φ. The new node-set is trivially recomputed; it is simply given by $V' = V - u$. Similarly, the set of measurements is given by $\mathbf{X}' = \mathbf{X} - \mathbf{x}_u$. Computation of the new edge-set E' is more complicated. The relational structure of the scene must be recomputed without the influence of the node u. For example, in the experiments detailed later in this section, the Voronoi

tessellation of the image plane and the associated Delaunay graph are recalculated when a node is dropped.

As before, we wish to optimise the MAP criterion for the matching configuration with respect to the unary measurements which are available. However in this situation we are interested in reconfiguring the graph structure in order to optimise the partition between consistently matched inliers and relational outliers. Formally the outliers are denoted by Φ. We must therefore optimise

$$P(f, \Phi|\mathbf{X}_f, \mathbf{X}_\Phi) = \frac{p(\mathbf{X}_f, \mathbf{X}_\Phi|f, \Phi)P(f, \Phi)}{p(\mathbf{X}_f, \mathbf{X}_\Phi)} \tag{4.5}$$

where $\mathbf{X}_f$ and $\mathbf{X}_\Phi$ are the unary measurements associated with the graph nodes and the null nodes respectively. In order to calculate the various terms in Equation 4.5 a model is required specifying the nature of the corruption present in the scene graph. Here we assume that noise elements are randomly distributed and uncorrelated with the true structure of the scene. As a consequence, the *a priori* probability of the null labels is independent of the match and, as in the development of our initial MAP criterion, we assume that the unary measurements on all nodes are conditionally independent. We may therefore factorise the conditional measurement density thus,

$$p(\mathbf{X}_f, \mathbf{X}_\Phi|f, \Phi) = \prod_{u \in G_1} p(\mathbf{x}_u|f(u)) \prod_{u \in \Phi} p(\mathbf{x}_u|\Phi) \tag{4.6}$$

assuming that the measurement probability densities on outliers are independent of the fact that they are in the null partition. Using Bayes theorem to re-write this equation in terms of *a posteriori* probabilities,

$$p(\mathbf{X}_f, \mathbf{X}_\Phi|f, \Phi) = \prod_{u \in G_1} P(f(u)|\mathbf{x}_u)\frac{p(\mathbf{x}_u)}{P(f(u))} \prod_{u \in \Phi} p(\mathbf{x}_u) \tag{4.7}$$

We now note that because the outliers are assumed to be *a priori* independent of the match we may factorise the joint probability thus;

$$P(f, \Phi) = P(f).P(\Phi) \tag{4.8}$$

Our model for the joint prior of the matching function, $P(f)$, remains unchanged, but whereas the joint probability of the match $P(f)$ is dependent on the structure of the scene, there is no meaningful structure to be found in the outliers. They are therefore independent of each other resulting in another factorisation;

$$P(\Phi) = \prod_{u \in \Phi} P(u \rightarrow \Phi) = P_{\phi}^{|\Phi|} \tag{4.9}$$

We now turn our attention to the optimisation of the partition between outliers and inliers. As with the matching phase, we are interested in locating the maximum *a posteriori* probability. The quantity of interest is therefore the MAP ratio between two states of the graph. The update procedure involves the deletion or reinstatement of image entities on a node-by-node basis. We need only evaluate the effect of reassigning one matched node to the null category. Under these conditions the ratio is

$$\frac{P(f_{G_1}, \Phi|\mathbf{X}_f, \mathbf{X}_\Phi)}{P(f_{G_1'}, \Phi|\mathbf{X}_f, \mathbf{X}_\Phi)} = \frac{P(f(u)|\mathbf{x}_u)}{P(u \in \Phi|\mathbf{x}_i).P(f(u))} \frac{P(f_{G_1})}{P(f_{G_1'})} \tag{4.10}$$

From this expression we can see that the assignment of a particular node is dependent on two factors; namely the ratio of joint matching prior and the ratio of the *a posteriori* matching probabilities for the current label and the null label. Both measurement information and graph structure of the nodes influences the classification process.

4.4.2 Relational Clustering

The structural effect of deleting a node is apparent in the change in the joint prior from $P(f_{G_1})$ to $P(f_{G_1'})$. In order to establish the influence of a single node reassignment we must compute the effect over all graph units structurally changed by the deletion. Our model of structural consistency is averaged over the cliques in the data-graph, and so we must examine those contributions that arise from modification of the cliques containing the node in question.

This set is constructed by identifying those nodes that form a clique with node u in graph G_1, i.e. $C_u - \{u\}$, and determining the new clique set for these nodes in the reconfigured graph G_1'. We let χ_u^+ denote the clique set of object u in graph G_1 and χ_u^- denote the

corresponding clique set in the reconfigured graph G_1'. In other words, χ_u^+ contains all the cliques which originally contained u as one of the external nodes; these are the graph units which are affected by the deletion of u. The set χ_u^- on the other hand contains the same cliques after the deletion of u and the associated reconfiguration of the relations in the graph. With this notation the change in the denominator of the MAP criterion caused by the deletion of the node u is proportional to

$$\Delta_u^- = P_\phi \sum_{j \in \chi_u^-} P(\Gamma_j) \tag{4.11}$$

By contrast, when considering the change in the numerator of the MAP criterion it is the clique set χ_u^+ to which we turn our attention. The corresponding change to the MAP criterion is proportional to

$$\Delta_u^+ = \frac{P[f(u)|\mathbf{x}_u]}{P[f(u)]} \sum_{j \in \chi_u^+} P(\Gamma_j) \tag{4.12}$$

With these two measures to hand, we can both delete and reinstate nodes in such a way as to monotonically increase the MAP ratio. We therefore delete node u provided $\Delta_u^+ < \Delta_u^-$ and reinstate the node if $\Delta_u^+ > \Delta_u^-$.

It should be noted that in assessing the change to the global MAP ratio we have confined our attention to the component of the average consistency criterion that is modified by node deletion. This effectively corresponds to ignoring the effect of the unmodified component which can be regarded as representing a constant pedestal consistency value. In order to justify this iterative reassignment approach, we will now illustrate that the modified consistency component is in fact proportional to the change in Kullback-Leibler entropy caused by node deletion. Consequently, the decision criteria specified in Equations 4.11 and 4.12 not only locate the global MAP, they also ensure that the reconstructed graph is the structure that maximises the Kullback-Leibler entropy with respect to the model.

To proceed, we note that the canonical way of assessing the impact of the change in a probability distribution evaluated over a set of discrete entities is to compute the Kullback–Leibler divergence (Kullback and Leibler, 1951; Kullback, 1987). Our strategy for assessing the relevance of nodes in the data graph is therefore to compute the change in Kullback-

Leibler entropy associated with reconfiguration of the graph and the consignment of nodes to the set of outliers Φ. In our application, we require a means of gauging the improvements associated with the deletion of nodes, which may be facilitated by comparing the matched relations from the original graph $\Gamma_u \in G_1$ with those in the modified graph $\Gamma'_u \in G'_1$. Adhering to our underlying philosophy of exploiting a relational model, we wish to gauge these differences in the light of the structure preserving relations residing in the dictionary, i.e. $S_i \in \mathcal{P}$. The Kullback-Leibler entropy which meets our requirements is

$$I_u(G_1, G'_1) = -\sum_{j \in \chi_u^+} \sum_{S_i \in \mathcal{P}} P(\Gamma_j|S_i) \ln \frac{P(\Gamma_j|S_i)}{P(\Gamma'_j|S_i)} \tag{4.13}$$

Substituting for $P(\Gamma_j|S_i)$ and exploiting the exponential nature of the probability distribution

$$I_u(G_1, G'_1) = -\sum_{j \in \chi_u^+} \sum_{S_i} \frac{K_{C_j}}{|\mathcal{P}|} \left\{k_e[\mathcal{H}(\Gamma_j, S_i) - \mathcal{H}(\Gamma'_j, S_i)] + k_s[\mathcal{S}(\Gamma_j, S_i) - \mathcal{S}(\Gamma'_j, S_i)]\right\} \times e^{-k_e\mathcal{H}(\Gamma_j,S_i) - k_s\mathcal{S}(\Gamma_j,S_i)} \tag{4.14}$$

The change in entropy associated with the iterative reassignment of the single node u is therefore proportional to the weighted change in Hamming distance and size difference over the set of modified cliques. This is a result which has a pleasing physical intuition. Our maximum entropy criterion for deleting a node is that it minimises the average weighted sum of Hamming distance and size difference between the relations in G_1 and G_2.

Now let us examine the change in the joint prior $P(f)$ during the deletion of a node. This change is given by

$$\begin{aligned} \Delta P(f) &= P(f_{G_1}) - P(f_{G'_1}) \\ &= \frac{1}{|V_1|} \sum_{j \in \chi_u^+} \frac{K_{C_j}}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} e^{-k_e\mathcal{H}(\Gamma_j,S_i) - k_s\mathcal{S}(\Gamma_j,S_i)} - e^{-k_e\mathcal{H}(\Gamma'_j,S_i) - k_s\mathcal{S}(\Gamma'_j,S_i)} \end{aligned} \tag{4.15}$$

This expression can be rewritten in the the following form which elucidates the relationship between the change in the joint prior and the Kullback-Leibler entropy:

$$\Delta P(f) = \frac{1}{|V_1|} \sum_{j \in \chi_u^+} \frac{K_{C_j}}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \left\{1 - e^{-k_e[\mathcal{H}(\Gamma'_j,S_i)-\mathcal{H}(\Gamma_j,S_i)]-k_s[\mathcal{S}(\Gamma'_j,S_i)-\mathcal{S}(\Gamma_j,S_i)]}\right\} \times e^{-k_e\mathcal{H}(\Gamma_j,S_i)-k_s\mathcal{S}(\Gamma_j,S_i)} \quad (4.16)$$

From Equation 4.16 it can be seen that that the change in the joint prior has a similar form to that of the Kullback-Leibler entropy; indeed under conditions of small changes in Hamming distance and size difference the exponential difference term can be approximated by a linear term. Since only one node is deleted in each modification of the graph, the Hamming distances and size differences under consideration generally change by a maximum of one, although because the graph structure is recomputed, the changes may be more than one. Simulation studies show that 90% of node deletions lead to a change in surrounding units of one or less nodes. If a linear approximation of the exponential is valid, the change in the criterion is given by

$$\Delta P(f) = \frac{1}{|V_1|} \sum_{j \in \chi_u^+} \frac{K_{C_j}}{|\mathcal{P}|} \sum_{S_i} \left\{k_e[\mathcal{H}(\Gamma'_j,S_i) - \mathcal{H}(\Gamma_j,S_i)] + k_s[\mathcal{S}(\Gamma'_j,S_i) - \mathcal{S}(\Gamma_j,S_i)]\right\} \times e^{-k_e\mathcal{H}(\Gamma_j,S_i)-k_s\mathcal{S}(\Gamma_j,S_i)} \quad (4.17)$$

This is proportional to the change in the Kullback-Leibler entropy. Configurations of optimum average consistency are therefore not only those of minimum average Hamming distance, they are also those that minimise the entropy of the relational model. Viewed as a relational clustering process, the maximum value of the joint prior is also of minimum entropy.

4.5 Experimental Comparison of Techniques

In this section we compare the relative performance of the three algorithms presented earlier. The performance evaluation is based on two factors, the fraction of correct matches (F_c) and the noise fraction (F_n). These are given by the formulas

F_c=*correct matches / maximum possible correct matches*

F_n=*incorrect matches / number of matches*.

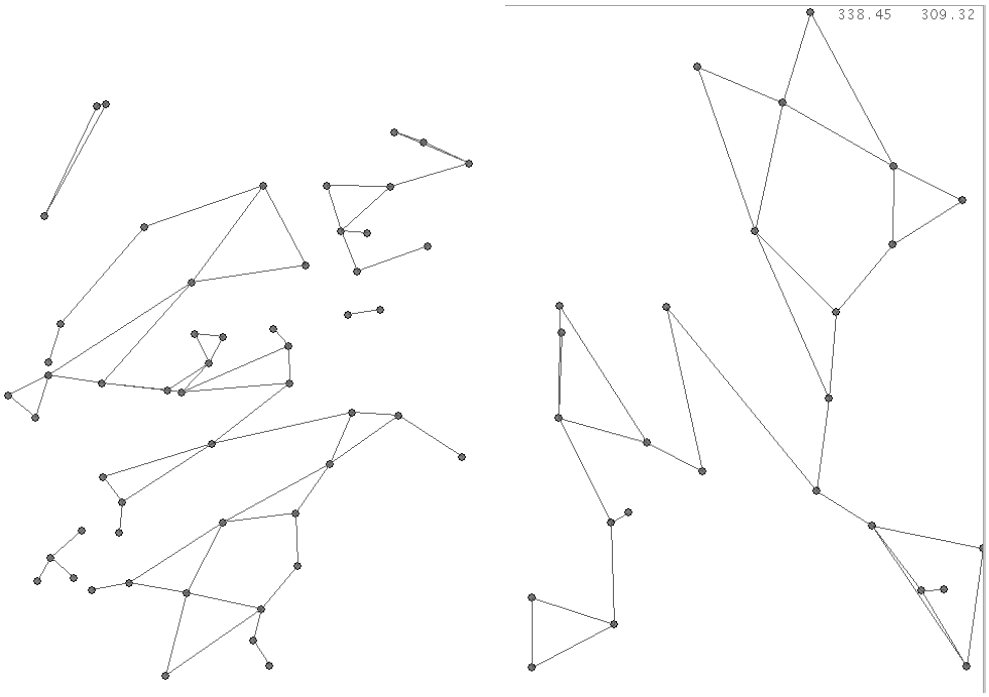


Figure 4.2: Corner graphs generated from the SAR and map data

With these definitions $0 \leq F_c \leq 1$ with 1.0 being the best performance, and $0 \leq F_n \leq 1$ with 0.0 being the best performance.

The first set of experiments is based on the SAR aerial data shown in Appendix A. We evaluate the matching performance on this data using two different sets of relations. The first set of relations are generated by the presence of corners between line segments. The second set are generated by a Voronoi tessellation seeded on the center points of the lines. These two relational abstractions generate very different edge densities; nodes under a Voronoi tessellation have an average of 5.5 connections, with up to 12 adjacent nodes for some. The corner relations have on average 2.4 neighbours with a maximum of 4. A corner graph representation of the data is shown in Figure 4.2

Table 4.1 summarises the results produced by the three algorithms derived in sections 4.2,4.3 and 4.4. From these results a clear difference can be seen, which is attributable to the different graph structures. On corner graphs constraint filtering is the most effective method. Both the optimisation and reconfiguration methods perform similarly poorly on this graph structure. However the situation is different when we come to study the Delaunay graph representation. The reconfiguration method improves dramatically, and

	Corner graph		Delaunay Graph	
	F_c	F_n	F_c	F_n
Constraint Filtering	0.71	0.35	0.96	0.56
Optimisation	0.52	0.78	0.57	0.77
Reconfiguration	0.43	0.81	0.77	0.47

Table 4.1: Summary of Results for noise removing schemes on real data

indeed it out-performs constraint filtering slightly in terms of the noise rejection figure. There is little difference in the performance of the optimisation scheme.

These differences can be understood in terms of the processes at work in the various approaches. The optimisation version of the null-labelling process compares unfavorably with the other methods; it is dogged both by problems of parameter control and a tendency to shuffle valid matches into the null match category from which they can no longer be successfully recovered. The graph reconfiguration method also performs poorly on the corner graphs. The nature of corner relations does not lend itself to reconstruction of the topology of the graph, since the dropping of a line segment leads only to the loss of the corner relations associated with that line; the rest of the graph is unaltered and no new relationships appear. There is little to be gained by attempting to reconfigure the relationships. In this case the algorithm is reduced to merely pruning out nodes of the graph.

With the Delaunay representation the graph is significantly altered by a node deletion; relationships disappear and new ones appear in their place. The reconfiguration technique is more potent since there are significant structural differences between the different graph configurations which the algorithm can use to determine the optimal partition between legitimate nodes and noise elements. Infact reconfiguration is marginally superior in terms of noise rejection to constraint filtering because of it's ability to restore the original uncorrupted relationships in the graph.

The second set of experiments involves the use of Voronoi triangulations of random point sets with controlled levels of corruption. These are described in more detail in Appendix A and are also used in Chapter 3. This allows a study of algorithm performance at varying levels of corruption. Figures 4.3 and 4.4 summarise the relative performance of the schemes

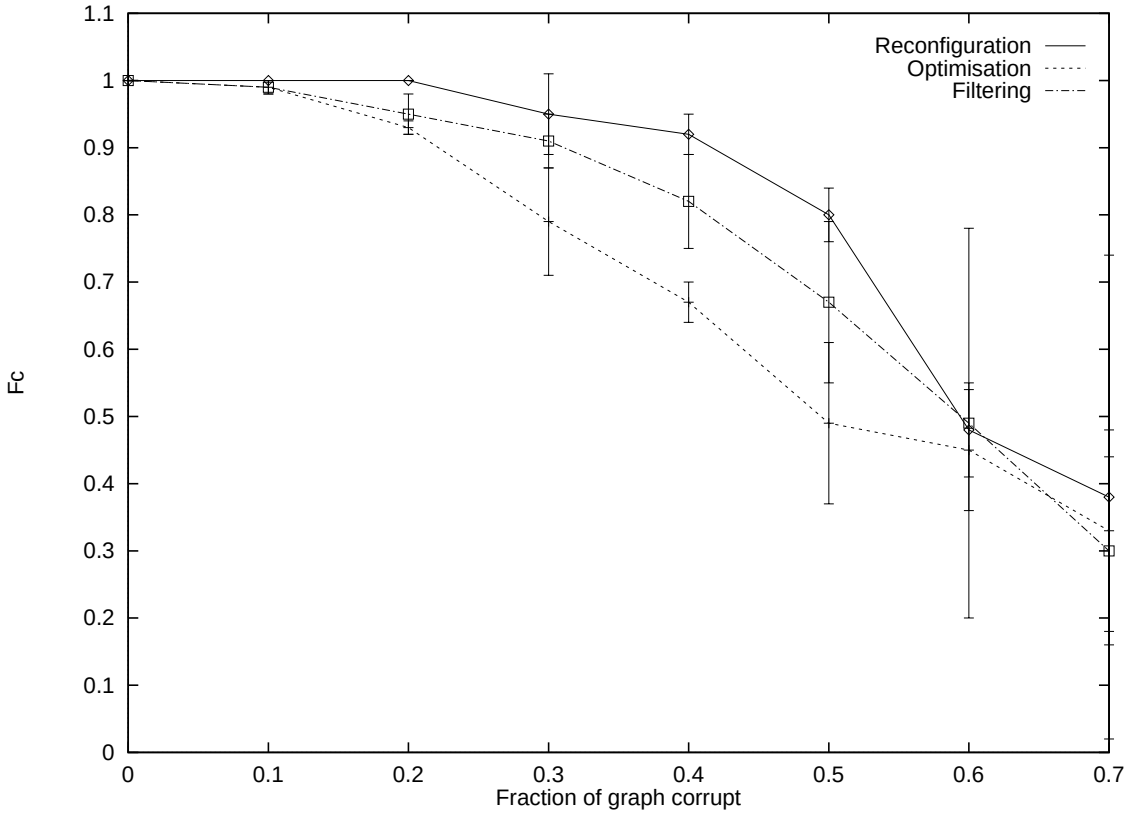


Figure 4.3: Correct Fraction for noise removing schemes

on this data.

These results lead to the conclusion that the optimisation approach is the inferior method of the three alternatives studied here. The algorithm is unable to allocate null labels during the matching phase without disturbing the patterns of correct matches.

A fuller study using Voronoi relationships (Figures 4.3 and 4.4) demonstrates that the graph reconfiguration method out-performs constraint filtering. There are two reasons for this performance advantage; firstly the reconfiguration approach is able to adjust the relations and restore the original graph structure after an element has been identified as noise. Secondly the constraint filtering algorithm uses a coarse measure of consistency when compared to our consistency criterion; rather than gauge the similarity between cliques, it relies on finding subgraph isomorphisms between the matched graphs. It is therefore less successful at setting the partition between noise and genuine matches. However in favour of constraint filtering is the fact that the approach takes minimal computational resources whereas graph reconfiguration requires the recalculation of relations typically

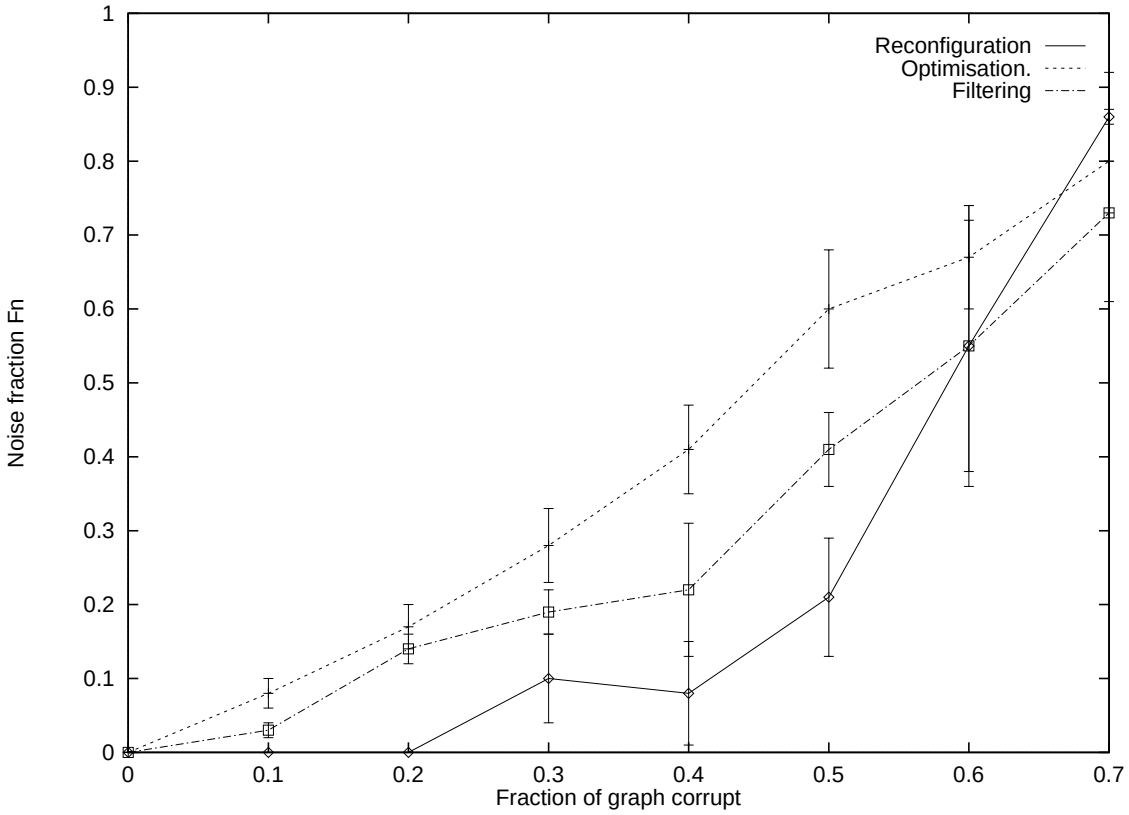


Figure 4.4: Noise Fraction for noise removing schemes

many thousands of times.

4.6 Conclusions

The following points summarize the finding in this chapter;

- A graph reconfiguration method has been developed which uses an objective consistency criterion to gauge the cost of graph edit operations. In the limit of small error and corruption probabilities the update operation also optimises the Kullback-Leibler entropy of the edited graph match.

- Results on both real-world data and simulated graphs demonstrate that, of the three methods studied here, the optimisation approach is significantly worse than the other two.

- The simulated graph study shows that the reconfiguration method is the most effective method for eliminating noise and recovering the correct match; however on the SAR

data, the constraint filtering approach recovers more of the correct match.

Chapter 5

Understanding the consistency Criterion

5.1 Introduction

We demonstrated in previous chapters, using experimental analysis, that the proposed consistency criterion is effective in matching graphs and eliminating noise and clutter from corrupted graphs. Because the criterion is based on a symbolic measure of consistency between graph sub-units, it is also possible to make some theoretical predictions. Armed with a model of the distribution of cliques in pattern space we can predict some properties of the criterion.

5.2 The Expectation Value of $P(f)$

In order to calculate the average value of the prior mapping probability, $P(f)$, we first require a pattern space model representing the distribution of structure-preserving mappings over the space of labellings. Since SPMs are in turn generated from the relations in the graph, this model is dependent on the type of relational structures present. The choice of these structures has an important role to play in determining the power of the constraints available to the relaxation algorithm. Here we develop a theory based specifically on Delaunay graphs, although the methodology is valid for other structures. Again, for the sake of simplicity, we

neglect the possibility of null labels.

We begin by writing the functional value explicitly in terms of the label error probability P_e thus

$$P(f) = \frac{1}{|V_1|} \sum_{C_i^{(1)}} \frac{1}{|\mathcal{P}|} \sum_{S \in \mathcal{P}} \prod P_e^{1-\delta(f(u),v)} (1 - P_e)^{\delta(f(u),v)} \tag{5.1}$$

Now consider the labelling on one particular clique match Γ_u with n nodes. With respect to this site, the dictionary $\mathcal{P}$ contains only one item which is fully congruent with Γ_u when it is correctly labelled. However because of both the overlap of SPMs and the need to enumerate all possible mappings of a clique, there will be a number of items which are partially congruent with Γ_u. The exact number and distribution of these items depends on the relational abstraction adopted when the graph is formed. In order to calculate the average value of $P(f)$ this distribution must be known.

Rather than evaluate the congruency of each clique with all it's SPMs, we make a simplification at this stage. We view the set of structure-preserving mappings at a particular clique ($S^{(u)}$) as having a 'principle' mapping which has the maximum possible congruency and a set of sub-mappings with lower congruency. Now provided that the number of sub-mappings with a congruency of one less than the maximum is small compared to $(1-P_e)/P_e$, then their influence on the functional will be negligible.

More formally, each clique in G_2 generates a set of SPMs comprising of the permutations of the non-centre nodes - we denote this set by $S^{(v)} = \{S_1^{(v)}, S_1^{(v)}, \ldots, S_k^{(v)}\}$. Now let

$$\epsilon_{min} = \min_{S \in S^{(v)}} \mathcal{H}(\Gamma_u, S) \tag{5.2}$$

be the minimum value of the Hamming distance for this set of mappings and let $N(\epsilon)$ be the number of mappings from $S^{(v)}$ which have ϵ as their Hamming distance from the clique Γ_u. The assumptions being made can be summarised by the following points:

- $N(\epsilon_{min}) = 1$
- $N(\epsilon_{min} + 1) << (1 - P_e)/P_e$
- $N(\epsilon_{min} + 2) << [(1 - P_e)/P_e]^2$

and so on.

If this condition is fulfilled only the principle mapping will be significant. The validity of the first assumption is dependent on the distribution of congruent labels and the way in which mappings are generated.

There are two processes which may violate the first assumption that there is only one principle mapping and we will deal with each in turn. The first process involves the generation of mappings in which some of the node pairings are the same. Each individual mapping is not independent of the others, rather they involve cyclic permutations and placements of dummy nodes. If there are no dummy nodes the similarity in the node pairings between mappings is confined to the centre node only and so the potential for the same minimum Hamming distance to appear in more than one mapping is small. As dummy nodes are added there is a much greater overlap between mappings due to the possibility of placing the dummy nodes in different positions within the mappings while retaining the rest of the node pairings. We must therefore confine the analysis to situations where dummy nodes are not a significant factor.

The other process we must consider is the case when the same minimum Hamming distance is generated by two mappings which have different node pairings, simply due to the chance congruency of labels. However if the Hamming distance is small enough to be significant influence on the criterion, the probability of it being repeated by the action of incorrectly congruent labels is small.

The other issue is that of the influence of mappings of larger Hamming distance than the minimum. This is not a significant if $N(\epsilon_{min}+1) << (1-Pe)/Pe, N(\epsilon_{min}+2) << [(1-Pe)/Pe]^2 \ldots$ The numbers of such mappings are difficult to predict but at arbitrarily small values of P_e this condition can be enforced. In practice, $P_e = 0.1$ gives $N(\epsilon) << 9N(\epsilon+1)$; without the action of dummy nodes the total number of mappings at each site rarely exceeds 9. With this consideration, we can expect this assumption to hold for say $P_e < 0.1$

As mentioned above, there is a chance that incorrect labels may accidentally be congruent with a SPM entirely unconnected with the correct mapping. The probability that two labels which have no place in the correct mapping are *not* the same is denoted by $P_e^{(w)}$.

We model these incorrect labellings by assuming a uniform distribution of labels when

a node is incorrectly labelled. As a consequence, the probability of a node u being wrongly labeled as node v in G_2 is given by $P_e/(V_2 - 1)$ and

$$P_e^{(w)} = 1 - \frac{P_e}{V_2 - 1} \tag{5.3}$$

We now partition the dictionary into a number of different sets, according to the potential congruency of each mapping, when correctly labelled, with Γ_u. For convenience we will denote each set by $\mathcal{P}_{n,m}$ where n is the number of potentially correct sites each mapping contains and m is the number of items in that set. Because of the overlap of cliques within the graph there will be a number of cliques $(2|C| - 2)$ which share two nodes with the central clique. The partition of the set of principle SPMs is therefore

$$\mathcal{P}_p = \mathcal{P}_{|C|,1} \cup \mathcal{P}_{2,2|C|-2)} \cup \mathcal{P}_{0,|V_2|-2|C|+1} \tag{5.4}$$

The probability of occurrence of a particular instance of a possible labelling is defined in the same fashion as our previous label distribution (Equation 2.10). For labels we expect to be correct, this is

$$P(f(u_k)|v_k) = \begin{cases} 1 - P_e & \text{if } f(u_k) = v_k \\ P_e & \text{if } f(u_k) \neq v_k \end{cases} \tag{5.5}$$

We also define a similar distribution for labels we would expect not to match

$$P(f(u_k)|v_k) = \begin{cases} 1 - P_e^{(w)} & \text{if } f(u_k) = v_k \\ P_e^{(w)} & \text{if } f(u_k) \neq v_k \end{cases} \tag{5.6}$$

Now let us examine the average value of the functional itself. Because of the average nature of the definition of the functional, the average value reduces to the average probability of one clique

$$< P(f) >=< \frac{1}{|\mathcal{P}|} \sum_{S \in \mathcal{P}} \prod_{u \in S} P_e^{1-\delta(f(u),v)} (1 - P_e)^{\delta(f(u),v)} > \tag{5.7}$$

and then expanding over the possible states of Γ_u denoted by $\omega \in \Omega$, the average value is given by

$$< P(f) >= \sum_{\omega \in \Omega} P(\Gamma_u = \omega) \frac{1}{|\mathcal{P}|} \sum \prod P_e^{1-\delta(f(u),v)} (1 - P_e)^{\delta(f(u),v)} \tag{5.8}$$

We may exchange the sum over all states of Γ_u with the product over labels in Γ_u to obtain

$$< P(f) >= \frac{1}{|\mathcal{P}|} \sum \prod \left\{ P[\delta(f(u), v) = 1](1 - P_e) + P[\delta(f(u), v) = 0] P_e \right\} \tag{5.9}$$

Armed with the pattern model defined in Equation 5.4 and with probability distributions for the various component labelling of this model, we can now calculate the average functional value to be

$$\begin{aligned} < P(f) > \quad = \quad & \frac{1}{|\mathcal{P}_p|} \Big\{ [(1 - P_e)^2 + P_e^2]^{|C|} \\ & + (2|C| - 2)[(1 - P_e)^2 + P_e^2]^2 \times [(1 - P_e)(1 - P_e)^{(w)} + P_e P_e^{(w)}]^{(|C|-2)} \\ & + (|V_2| - 2|C| + 1)[(1 - P_e)(1 - P_e)^{(w)} + P_e P_e^{(w)}]^{|C|} \Big\} \end{aligned} \tag{5.10}$$

This expression can be considerably simplified by the substituting for the weights associated with correct and incorrect labels, i.e. $\alpha_c = (1 - P_e)^2 + P_e^2$ and $\alpha_w = (1 - P_e)(1 - P_e^{(w)}) + P_e P_e^{(w)}$. The average functional value therefore becomes

$$< P(f) >= \frac{1}{|\mathcal{P}_p|} \left\{ \alpha_c^{|C|} + (2|C| - 2)\alpha_c^2 \alpha_w^{(|C|-2)} + (|V_2| - 2|C| + 1)\alpha_w^{|C|} \right\} \tag{5.11}$$

Although Equation 5.11 has been developed with specific reference to Delaunay graphs, study of the form of the expression reveals a more general applicability to graph structures. If the mapping items generated by the graph can be partitioned in a similar fashion to that described in Equation 5.4 then an equivalent expression for the average functional value applies. Formally, we partition the principle SPMs as follows

$$\mathcal{P}_p = \bigcup_{\mathcal{P}_{n,m} \in \mathcal{P}_p} \mathcal{P}_{m,n} \tag{5.12}$$

The average value of the functional is now simply given by

$$< P(f) >= \frac{1}{|\mathcal{P}_p|} \sum_{\mathcal{P}_{n,m} \in \mathcal{P}_p} m \alpha_c^n \alpha_w^{(|C|-n)} \tag{5.13}$$

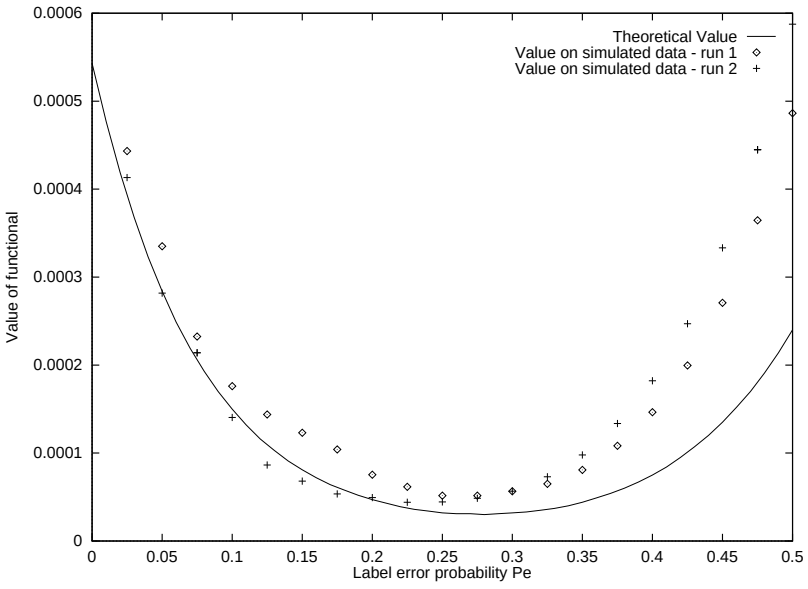


Figure 5.1: Comparison of true and predicted functional values

5.2.1 Experimental Validation

It now remains to compare this expression with the value calculated from experimental runs of the matching algorithm. Figure 5.1 shows a graph of the functional value sampled over the course of two matching runs. The solid line represents the theoretically predicted value of the functional. Both are plotted against the true label error probability which is generated by a 'master' matching which gives the true labelling of the nodes.

Figure 5.1 shows that our pattern space model leads to an accurate model of the processes at work in the matching process at lower levels of P_e. The experimental and theoretical values begin to diverge towards $P_e = 0.5$, in keeping with the assumptions made earlier.

It is important to note that the values in Figure 5.1 are compared with a label error probability which was generated with reference to a 'master' match which needs to be gained from some other source. Clearly in realistic applications no such information is available. However knowledge of the expectation value of the functional provides a method by which we can determine the prevailing label error probability.

Since we do not know what the value of the label error probability is, we commence by allowing the value used in the calculation of the functional to vary. The true level of errors remains unchanged. If the value used by the algorithm is denoted Q_e, we have

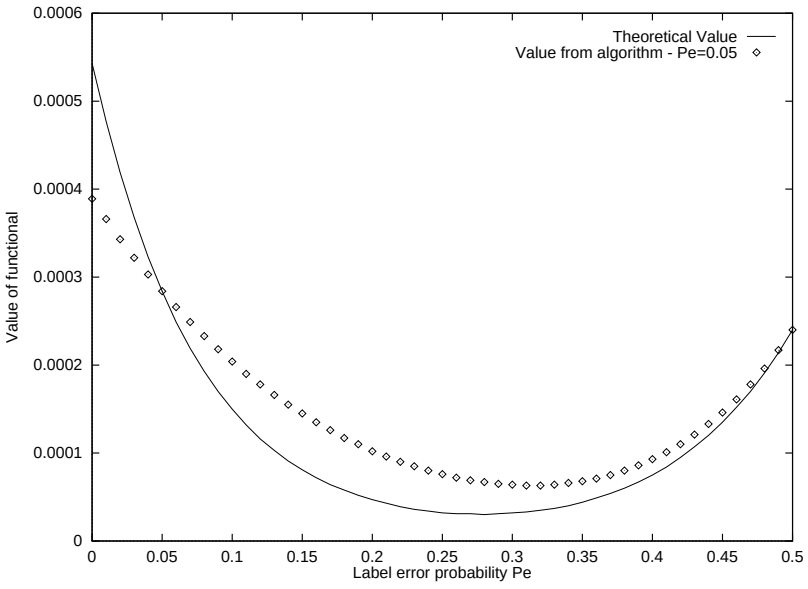


Figure 5.2: Finding the true value of Pe

$$\alpha_c' = (1 - Q_e)(1 - P_e) + Q_e P_e$$

$$\alpha_w' = (1 - Q_e)(1 - P_e^{(w)}) + Q_e P_e^{(w)}$$

These values are then used to calculate a new average functional $\mathcal{F}'$. This is the functional value that the matching algorithm will return. Let us now examine a plot of both $< \mathcal{F}' >$ and $< \mathcal{F} >$ (see Figure 5.2). The prevailing value of P_e can then be determined by the intersection point of the two curves.

5.3 Stochastic Relaxation

Stochastic relaxation has often been proposed as a suitable method for escaping from local optima of a functional. In particular Geman and Geman (Geman and Geman, 1984) originated a method of stochastic relaxation based on an analogy with statistical physics. They compare the labelling of objects in a Markov Random Field with annealing in a physical system. Label updating is performed in a non-deterministic fashion which effectively allows transitions between the current state and a random new configuration with a probability equal to the ratio of the probabilities of the new state Ω' and the previous configuration Ω,

i.e. the quantity

$$q = \frac{P(\Omega')}{P(\Omega)} = e^{-\beta[\epsilon(\Omega')-\epsilon(\Omega)]} \tag{5.14}$$

is calculated; the transition is allowed with probability q (or 1 if $q > 1$). In other words, labelling changes which increase $\epsilon(\Omega)$ are allowed with a probability q, whereas changes which decrease $\epsilon(\Omega)$ are always allowed. This has the effect of producing new states Ω in proportion to their probability $P(\Omega)$ when ϵ is the Gibbs energy.

The probability distributions used by Geman and Geman are based on Gibbs distributions containing the parameter $T = 1/\beta$, equivalent to the temperature in a physical system. They show that by reducing the temperature in a particular fashion, namely at iteration k, $T_k \geq \frac{c}{\log(1+k)}$ local minima of the energy function can be avoided.

In order to relate the previously described discrete relaxation scheme to this statistical physics analogy, an equivalent Boltzmann distribution is required to represent the system. We begin by hypothesising a set of possible states, denoted by $\Pi = \{\pi_1, \pi_2, \ldots\}$, for the clique under match, Γ. These states are the equivalent ingredients in the statistical mechanics framework to the SPMs in the discrete relaxation scenario. There is one state for each SPM, and if the match is in state π this is equivalent to matching Γ to the corresponding SPM.

The clique matching probability in a given state π can then be calculated and is given by

$$P(\Gamma, \pi) = \frac{K}{|\Pi|} e^{-k\mathcal{H}(\Gamma,\pi)} \tag{5.15}$$

In drawing an analogy with statistical physics, we compare this expression with the partition function for a system of particles with possible configurations $\Phi = \{\phi_1, \phi_2, \ldots\}$. This partition function is given by

$$Z = \sum_{\phi \in \Phi} e^{-\beta \epsilon_\phi} \tag{5.16}$$

We therefore define the equivalent partition function for the discrete relaxation system to be

$$Z = \frac{K}{|\Pi|} \sum_{\pi \in \Pi} e^{-k\mathcal{H}(\Gamma,\pi)} \tag{5.17}$$

From this starting point a number of statistical and 'physical' properties of the system can be calculated. For example, the equilibrium Gibbs potential of a particular clique can

easily be derived from the formula

$$U(\Gamma) = -\frac{1}{Z}\frac{\partial Z}{\partial k} \tag{5.18}$$

Combining Equation 5.17 and Equation 5.17 we obtain

$$U(\Gamma) = \frac{\sum_{\pi} \mathcal{H}(\Gamma,\pi)e^{-k\mathcal{H}(\Gamma,\pi)}}{\sum_{\pi} e^{-k\mathcal{H}(\Gamma,\pi)}} \tag{5.19}$$

This expression is worthy of further comment. The effective potential for the clique mapping Γ is the average value of the Hamming distance with respect to the available SPMs. Examination of the individual terms in the numerator of expression 5.19 reveals an important feature of the Gibbs energy; as $\mathcal{H} \rightarrow \infty$, $\mathcal{H}e^{-k\mathcal{H}} \rightarrow 0$ and when $\mathcal{H} = 0$ then $\mathcal{H}e^{-k\mathcal{H}} = 0$. The exponential weighting of Hamming distance suppresses the influence of items of large Hamming distance, while items of small Hamming distance also have little contribution. The main contribution to the potential is therefore from items of intermediate values of Hamming distance. Figure 5.3 shows this factor graphically. This is an important feature in terms of the algorithm's ability to handle a large dictionary of SPMs; items with a large Hamming distance of order $|C_{\Gamma}|$ which have little influence on the Gibbs energy correspond in general to mappings which in reality have no connection with the correct match. Items with zero Hamming distance will correspond to mappings in which a consistent match has already been found. Attention is naturally focussed on the intermediate items on which consistency still needs to be imposed. This structure limits the effects of inter-pattern competition which dog linear systems, and has been shown to lead to a vastly improved storage capacity and enhanced pattern reconstruction abilities over the Hopfield network in the case of binary memories (Hancock and Kittler, 1993; Gardner, 1986).

5.3.1 The Entropy of Matching

From the same starting point we can also calculate another property of the system, the thermodynamic entropy. The entropy of an individual clique with respect to the matching state π is given by

$$S = k' \ln P(\Gamma,\pi) \tag{5.20}$$

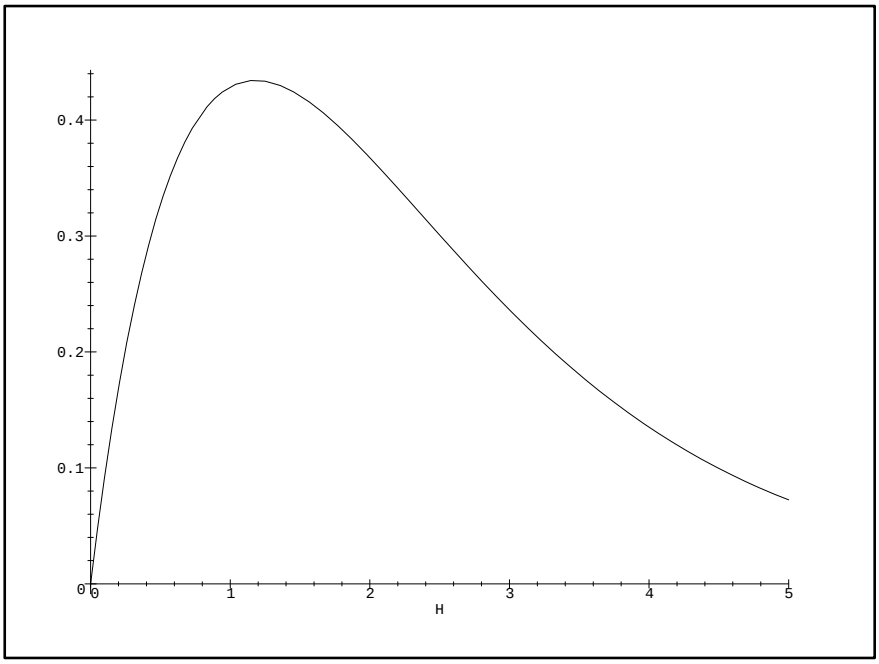


Figure 5.3: Contribution to Gibbs potential from varying Hamming distances

Substituting for the expression for $P(\Gamma, \pi)$ (Equation 5.15), we obtain

$$S(\Gamma, \pi) = k' \ln \left[\frac{K}{\Pi} e^{-k\mathcal{H}(\Gamma,\pi)} \right] \tag{5.21}$$

The average value of the clique entropy is given by averaging over the possible states of the match thus $< S >= \sum_\pi S(\Gamma, \pi) P(\Gamma, \pi)$. Consequently, the average entropy is:

$$< S >= \frac{K k'}{|\Pi|} \sum_{\pi \in \Pi} [\ln(K/|\Pi|) - k\mathcal{H}(\Gamma, \pi)] e^{-k\mathcal{H}(\Gamma,\pi)} \tag{5.22}$$

5.3.2 Equivalent Boltzmann Distribution

As we have seen, the probabilities which are available model the Boltzmann distributions for the different dictionary items in the set of SPMs. With these distributions we can stochastically update the *state* π. This corresponds to optimal selection of the dictionary item which corresponds to a particular fixed labelling Γ. This is of little interest as far a improving the quality of match is concerned. Rather we need to optimise the clique labelling itself. In order to do this an equivalent Boltzmann distribution for the labelling is required, which we denote $Q(\Gamma)$. As required by the theorem of total probability, we can expand Q over the possible states of the system, i.e.

$$Q(\Gamma) = \sum_{\pi \in \Pi} Q(\Gamma, \pi) \tag{5.23}$$

We now require a method of relating the joint probabilities $Q(\Gamma, \pi)$ to the original joint probabilities of the different states. Following (Cross et al., 1995), we use a result from information theory, namely that the probability distribution which best models a known distribution can be found by minimising the Kullback-Leibler divergence. The divergence between the conditional matching probabilities $P(\Gamma|\pi)$ defined over the state-space of the dictionary and the new conditional probabilities of the Boltzmann distribution $Q(\Gamma|\pi)$ is given by

$$\Delta(\Gamma) = \sum_{\pi \in \Pi} P(\Gamma|\pi) \ln \frac{P(\Gamma|\pi)}{Q(\Gamma|\pi)} \tag{5.24}$$

Before continuing, we must apply additional constraints to Q in order to provide a unique solution for the minimal point of the Kullback-Leibler divergence. We note that we have no interest in which state is involved in the optimal matching; the Boltzmann distribution we are looking for is a function of the labelling only. Consequently we assume

that the probabilities q are independent of the state of the match, i.e. that $Q(\Gamma, \pi_1) = Q(\Gamma, \pi_2) = Q(\Gamma, \pi_m) \forall m$. Under this assumption we obtain $Q(\Gamma) = \sum_\pi q(\Gamma|\pi)p(\pi) = Q(\Gamma|\pi)$. Substituting the expression for the original conditional matching probabilities, we obtain

$$\Delta(\Gamma) = K \sum_\pi \Big[\ln K - k\mathcal{H}(\Gamma, \pi) - \ln Q(\Gamma|\pi) \Big] e^{-k\mathcal{H}(\Gamma,\pi)} \tag{5.25}$$

The Kullback-Leibler divergence is minimal when the distributions are identical, i.e. $\Delta(\Gamma) = 0$. Using this condition;

$$\ln K \sum_\pi e^{-k\mathcal{H}(\Gamma,\pi)} - k \sum_\pi \mathcal{H}(\Gamma, \pi) e^{-k\mathcal{H}(\Gamma,\pi)} - [\ln Q(\Gamma)] \sum_\pi e^{-k\mathcal{H}(\Gamma,\pi)} = 0$$

and therefore the equivalent Boltzmann distribution for the clique is

$$Q(\Gamma) = K e^{-kU(\Gamma)} \tag{5.26}$$

Hence $U(\Gamma)$ is the equilibrium Gibbs potential for the equivalent Boltzmann distribution of clique Γ. It is equivalent to the set of distributions over the state-space of mappings in the sense that the Kullback-Leibler entropy between the two is zero. We are now justified in performing simulated annealing on the potential $U(\Gamma)$.

5.4 Conclusions

In this chapter we have predicted the average value of the matching criterion under some limiting assumptions about the nature of the pattern space and parameter values, which confine the validity of the expression to small values of the label error probability Pe. This expression was show to be an accurate prediction of the criterion experimentally, for values of $P_e < 0.2$.

We then drew on the partition function of the system to calculate the Gibbs energy of the labelling with respect to the set of SPM's. The Gibbs energy was simply the average Hamming distance over the set of SPM's. Further analysis demonstrated that the equivalent Boltzmann distribution for the clique labelling was an exponential function of this energy.

Chapter 6

A Comparative Study of Optimisation-based Inexact Matching Schemes

6.1 Introduction

In the preceding chapters we have developed a compound exponential criterion for performing graph matching tasks which develops earlier ideas of Hancock and Kittler (Hancock and Kittler, 1990a) who introduced the concept of a label error process and applied it to low-level labelling problems. In the past little attention has been directed at the mathematical form of the criterion for matching tasks. In general authors have been little concerned with how the choice of criterion affects the fidelity and robustness to noise of matching algorithms. Instead they have opted for simple expressions which permit ease of computation and analysis. Examples of this can be seen in (Faugeras and Berthod, 1981; Shapiro and Haralick, 1981; Boyer and Kak, 1988). In the first part of this chapter, we examine some alternative forms of the matching criterion, show how they relate to the exponential approach described in previous chapters and discuss the robustness of different methods.

Infact it has been the optimisation process that has interested many authors, with elaborate update procedures (Geman and Geman, 1984; Lloyd, 1983) or search techniques (Shapiro and Haralick, 1981; Messmer and Bunke, 1994) overcoming problems of local op-

tima in the matching function. One example of this is in Geman and Geman's work (Geman and Geman, 1984). Here they develop an elaborate stochastic update procedure known as simulated annealing to escape from local minima of the matching function. However it may be that the need for stochastic optimisation methods is largely inflicted by the choice of criterion. In the final part of this chapter we assess the need for a more elaborate optimisation scheme.

6.2 The Linear Approximation

Traditionally the relaxation technique has involved the evaluation of a labelling energy or criterion which is a linear sum of error terms, with each error term being linear or quadratic in terms of the adopted distance measure. For example, it is precisely the quadratic form of the consistency criterion which is optimised by the Hummel and Zucker (Hummel and Zucker, 1983) probabilistic relaxation scheme. Linear compatibility models are common place in the literature on relaxation labelling (Bhanu and Faugeras, 1984; Izumi et al., 1992; Lloyd, 1983; Ranganath and Chipman, 1992; Ton and Jain, 1989).

However recent efforts have established a fundamental weakness in this technique, particularly in the context of pattern recognition problems. This is highlighted particularly by the Hopfield network (Hopfield, 1984) which essentially operates using a linear criterion. It has been established (Gardner, 1986) that the Hopfield network is severely limited in the number of patterns it can store and recognize; the limitation stems from problems of inter-pattern competition. For a large number of patterns, the contribution from non-matching patterns swamps the signal from the one matching pattern.

In contrast, recent investigations (Hancock and Kittler, 1993) have shown that the sum of exponential inter-pattern distances is superior in its ability to cope with large sets of patterns. The exponential element to the matching function effectively suppresses the contribution from outlier patterns, which in the linear case can dominate the criterion. In this sense the linear expression is non-robust to pattern noise since an arbitrarily placed outlier can arbitrarily change the value of the criterion.

In this section we establish a linear approximation to our global criterion of match,

and use this as a point of comparison with other such techniques. We also investigate the performance of this linear cost function.

6.2.1 An Approximate Criterion

We begin by examining the expression for the clique matching probability:

$$P(\Gamma) = \frac{K_C}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \exp[-k_e \mathcal{H}(\Gamma, S_i) - k_\phi \mathcal{S}(\Gamma, S_i)] \tag{6.1}$$

We are interested in computing a linear approximation to the exponential appearing in the above expression using the Taylor expansion. This expansion is valid when $k_e \mathcal{H}(\Gamma, S) \rightarrow 0$ and $k_\phi \mathcal{S}(\Gamma, S) \rightarrow 0$. This limit occurs when the corresponding constants k_e and k_ϕ approach zero, i.e when the error probability $P_e \simeq \frac{1}{2}$ and the node loss probability $P_\phi \simeq \frac{1}{2}$. Under these conditions the exponentials of in the summation in equation 5 can be expanded in terms of a Taylor power series,

$$P(\Gamma) = \frac{K_C}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \left\{1 - k_e \mathcal{H} - k_\phi \mathcal{S} + \frac{1}{2}[k_e \mathcal{H} + k_\phi \mathcal{S}]^2 \ldots \right\} \tag{6.2}$$

Limiting the expansion to linear terms, we obtain

$$P(\Gamma) = K_C \left\{1 - \frac{1}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \left[k_e \mathcal{H} + k_\phi \mathcal{S}\right] \right\} \tag{6.3}$$

The resulting global consistency criterion under consideration is then

$$\mathcal{F}_h = \frac{1}{|V_1|} \sum_{C_j^{(1)} \subset V_1} K_C \left\{1 - \frac{1}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} [k_e \mathcal{H}(\Gamma, S_i) + k_\phi \mathcal{S}(\Gamma, S_i)] \right\} \tag{6.4}$$

Therefore, in the situation when clique size differences are not important (if, for example we confine ourselves to pair units only, or any units of the same cardinality) the global criterion may be approximated by the sum of Hamming distances to the SPMs. The reader should note however that this approximation is only accurate when $P_e \simeq \frac{1}{2}$ and there is a clear divergence between the two schemes as P_e is reduced.

Several features of this approximation deserve further comment. In the first instance, the resulting relaxation scheme minimises the total Hamming distance to the set of structure preserving mappings. This minimisation of a linear function of Hamming distance as a

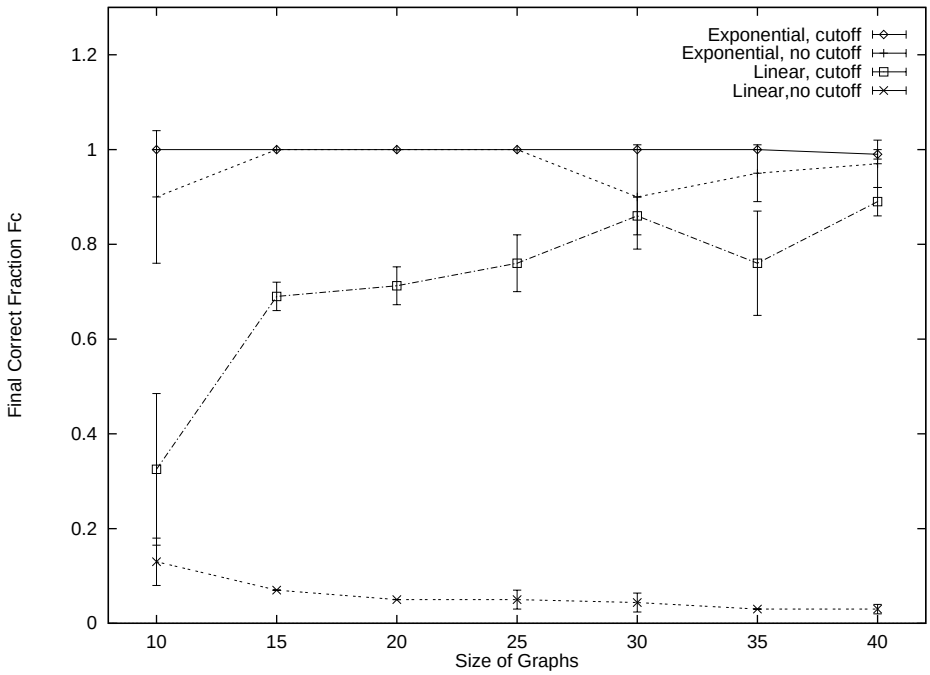


Figure 6.1: Performance of linear and exponential forms over a range of graph sizes

measure of total error is precisely the function performed by the Hopfield memory in the associative recovery of binary patterns. We can anticipate that the linear approximation will suffer in exactly the same way as the Hopfield memory. In particular we would expect the criterion to be limited in its capacity to accomodate a large set of SPM's. As the size of the model graph increases, the number of SPMs generated from it rises and the matchability should decrease. Infact, the Hopfield memory can distinguish $0.14N$ patterns (Gardner, 1986) where N is the number of nodes in the pattern; for the graph matching problem on planar graphs there are a minimum of $N - 1$ patterns generated from each graph clique.

6.2.2 Experimental Comparison

Theory suggests that the linear approximation will become increasingly poor as the size of graphs to be matched increases. In this section we present experiments comparing the behaviour of the linear scheme with the exponential formulation over a range of graph sizes. We also compare the performance on graphs at different levels of corruption.

The clearest way of illustrating the effect of increasing the number of mappings explored is to compare the performance of the linear algorithm over both the full set of SPMs and

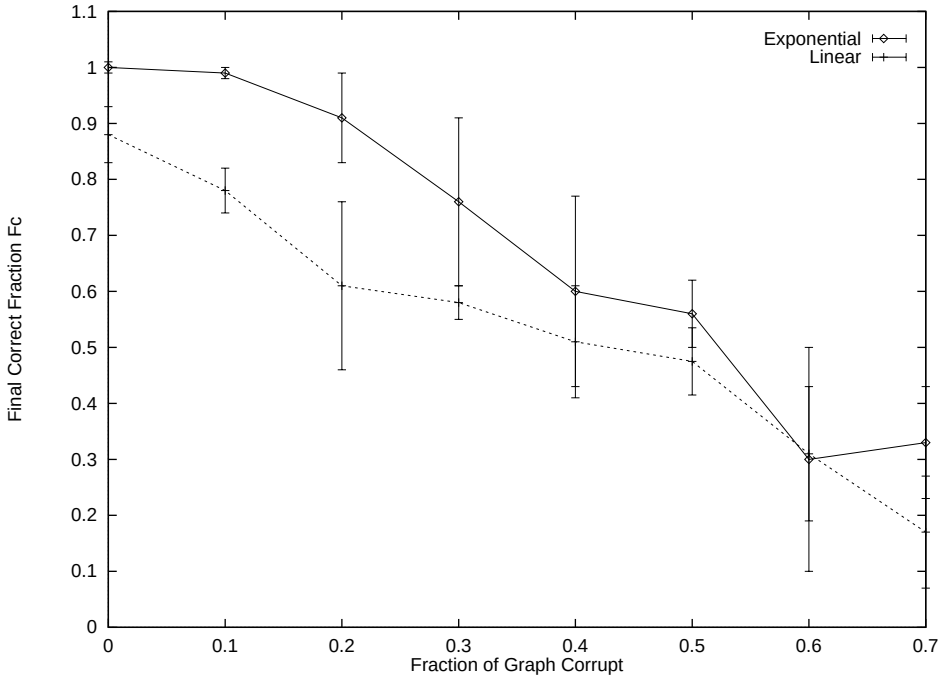


Figure 6.2: Performance of linear and exponential forms under corruption

over a limited number of SPMs. We can limit the number of mappings explored by applying a threshold to the initial probabilities of node mappings. Consider the case when we are trying to evaluate the probability of a particular clique, ie $P(\Gamma) = \sum P(\Gamma|S)P(S)$. We can assume that if the initial probability $P(f(u_\Gamma) = v_S)$ is very small (where u_Γ is the centre node of clique Γ and v_S is the centre node of S), that we may ignore the possibility that S is the matching clique to Γ. In this case, we do not evaluate contribution from the mapping S. The number of mappings being evaluated is drastically reduced, a fact which we would expect to benefit the linear criterion. Figure 6.1 illustrates the effect of applying this cut-off; the difference to the linear scheme is clear - without the cut-off the linear scheme produces results commensurate with a random labelling (one node correct). When the cut-off is applied, this increases to around a 90% correct labelling. The drop-off in the performance of the cut-off linear criterion for smaller graphs can be attributed to the fashion in which the cut-off is generated; for small graphs the initial probabilities of individual nodes are larger and the cut-off removes a smaller proportion of mappings. Finally, the figure shows that the cut-off has little effect on the full exponential criterion.

Figure 6.2 illustrates the relative performance of the linear and exponential schemes

with the cut-off applied over a range of corruption levels. The exponential scheme clearly out-performs the linear scheme over all values of corruption up to 50%. The linear scheme starts at around 90% of nodes correctly matched with no graph corruption and begins to drop away immediately. This is in contrast to the exponential approach which results in much higher levels of performance.

6.3 Squared Error Criteria

Another common approach is to use a mean square error criterion to measure the divergence between structural units. The main motivation for this approach is related to the well understood properties of mean square error functions, and their foundation in statistical estimation procedures. In the case where the deviations are Gaussian, minimising the mean square error correspondingly maximises the log likelyhood, i.e. when $P(\alpha) = \prod_i (1/\sigma\sqrt{2\pi}) \exp[-e_i^2/2\sigma^2]$, minimising the total squared error $\sum_i e^2$ maximises $P(\alpha)$. They are not, however, robust when estimating noisy data (Kittler and Taylor, 1994) because they suffer from one flaw; all the samples are assumed to be part of the inlier distribution. This is rarely the case in a realistic problem. Infact, for the quadratic criterion large pattern deviations which are associated with outliers produce a larger contribution to the criterion than smaller deviations. The criterion can therefore be dominated by outlier patterns which have no connection to the true pattern. In contrast the exponential criterion gives very little weight to large errors.

It is interesting at this point to compare our exponential criterion with that used by Boyer and Kak (Boyer and Kak, 1988) and Sarkar and Boyer (Sarkar and Boyer, 1993). Their criterion of match is based on the entropy of inter-primitive distances of objects and relations in the graph. These inter-primitive distances are founded on measurement information relating to these primitives. We can write a simplified version of their criterion as

$$\mathcal{F}_{bk} = \sum_{i \in V_1} \sum_{j \in V_2} \log d(p_i, p_j) \tag{6.5}$$

If we were to adopt a Gaussian model of attribute deviations, ie $d(p_i, p_j) = exp[(\mathbf{x}_i - \mathbf{x}_j)\Sigma^{-1}(\mathbf{x}_i - \mathbf{x}_j)^T]$ this expression becomes the mean square error criterion discussed earlier:

$$\mathcal{F}_{bk} = \sum_{i \in V_1} \sum_{j \in V_2} (\mathbf{x}_i - \mathbf{x}_j)\Sigma^{-1}(\mathbf{x}_i - \mathbf{x}_j)^T \tag{6.6}$$

A point of contact can be made with the exponential criterion if we were to adopt the Hamming between primitives as our measure of relational unit similarity, as opposed to the attribute deviations measured by the Mahalanobis distance. With this modification, we can see that under the assumption of Gaussian distributions, the Boyer and Kak approach is equivalent to the quadratic approximation to the exponential Hamming distance scheme. In other words if we make the substitution $\mathcal{H}(\Gamma, S) \rightarrow (\mathbf{x}_i - \mathbf{x}_j)\Sigma^{-1}(\mathbf{x}_i - \mathbf{x}_j)^T$ then we can see that the case of the Gaussian Boyer and Kak functional is equivalent to a quadratic approximation of the exponential function, except that unit similarity is measured in terms of attribute measurement deviations rather than the symbolic Hamming distance.

6.3.1 Experimental Comparison

From the analysis in the preceding section we anticipate that the mean square error should be even less robust to relational noise than the linear approximation, because outlier patterns are given greater weight than inliers. Here we analyse the performance of all three schemes (the linear, square and exponential schemes) over a range of levels of corruption.

It is clear from Figure 6.3 that little improvement is gained at all by applying the mean-square-error criterion; as anticipated it is not at all robust to outliers.

6.4 The Hard Limit

We begin this approximation by eliminating the effect of null-labels; this corresponds to setting the null-relation probability $P_\phi = \frac{1}{2}$. In this mode we allow comparison between cliques of different sizes with no penalty. Alternatively the size difference term will disappear if we operate on structural units of the same size. Under these conditions $k_\phi = 0$, and the clique probability is given by

$$P(\Gamma_j) = \frac{K_C}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \exp[-k_e \mathcal{H}(\Gamma_j, S_i)] \tag{6.7}$$

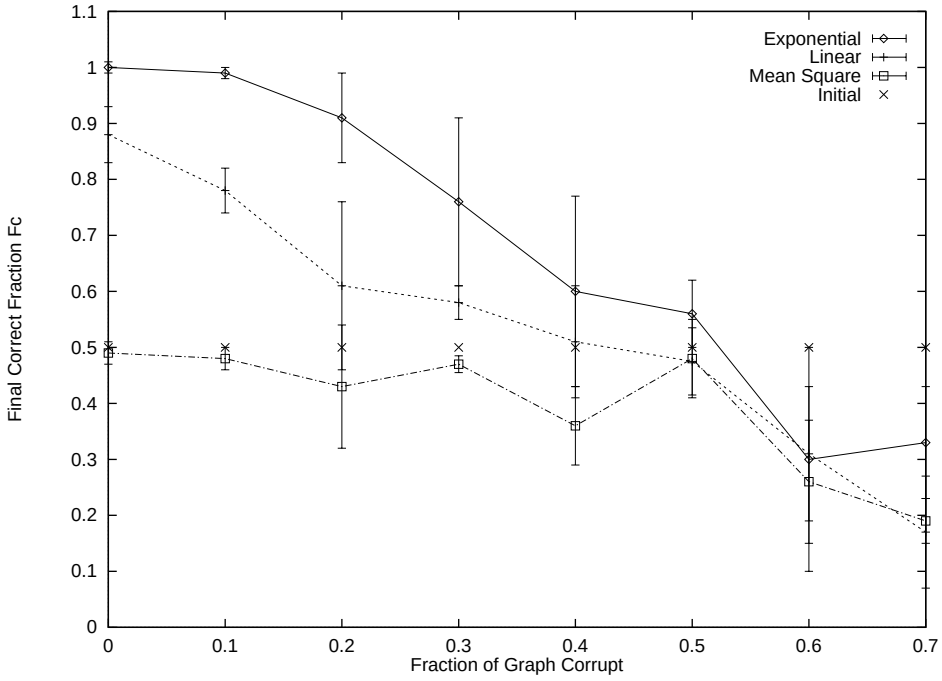


Figure 6.3: Performance of square, linear and exponential schemes under corruption

To gain some computational advantage, we may wish to eliminate the exponentials appearing in equation 6.7 and thereby simplify the calculation of probability. This can be achieved by examining the case when the label error probability becomes vanishingly small, i.e. $P_e \rightarrow 0$. Under these conditions the exponentials approach their delta function limits.

$$P(\Gamma_j) \simeq \frac{K_C}{|\mathcal{P}|} \sum_{S_i \in \mathcal{P}} \delta[\mathcal{H}(\Gamma_j, S_i)] \tag{6.8}$$

Each clique now contributes to the global consistency measure in a binary fashion; completely consistent cliques contribute an amount $\frac{K_C}{|\mathcal{P}|}$, while everything else contributes zero. The consistently matched portion of the graph contains those cliques for which a zero Hamming distance exists in the set of allowed mappings.

The delta-function form of the criterion has significant computational advantages; the evaluation of equation 6.8 can be implemented by simple table lookup. Once a clique has reached a zero Hamming distance configuration, there is no further improvement to be gained by label updating. Such a clique may be removed as a candidate for further reconfiguration, leaving the set $V - \Lambda$ only under consideration. Computational resources can be focussed on cliques adjacent to nodes in Λ, allowing consistency to spread as a

"brushfire" from the consistent portions of the graph. This technique obviates the need for an exhaustive iterative search over all labels and sites.

However, despite the computational advantages, it is immediately obvious that the resulting consistency measure will lead to a deadlocked updating procedure with even moderate departures from consistency. If the full consistency of individual cliques cannot be restored by a single label update, the update process will be unable to improve consistency by deterministic means. The only way to escape these deadlocks is via an expensive stochastic optimisation scheme.

Shapiro and Haralick (Shapiro and Haralick, 1981) propose a measure of graph similarity which is, in essence, the same as this process. They effectively count the number of consistently matched relations, whilst allowing the addition of null nodes at no additional cost. The Shapiro and Haralick scheme minimises the numbers of inconsistent mappings using a search method. It is equivalent to the hard limit of the discrete relaxation method in terms of the cost function when cliques are used as relations. Their strategy also encounters problems of deadlock which they resolve with a forward checking and backtracking search algorithm rather than stochastic optimisation.

6.4.1 Experimental Comparison

The experimental evaluation presented in Figure 6.4 compares the performance of this scheme with that of the full exponential scheme. The comparison is made using Delaunay graphs representing the structure of random dot patterns, with a range of levels of corruption. It is clear from these results that, as anticipated, the hard approximation performs poorly at all levels of relational corruption. At moderate levels, the full exponential scheme significantly out-performs the simpler alternative. However at high levels of corruption both versions fare equally badly with neither significantly improving the initial labelling.

Under circumstances of moderate levels of corruption the hard limit of delta functions can therefore be inappropriate. Partially matched cliques have no influence in the consistency measure, causing poor matches to deadlock the scheme. We can further clarify how the delta-function scheme operates by studying the case of initialisation errors only with no graph corruption. Figure 6.5 illustrates this comparison.

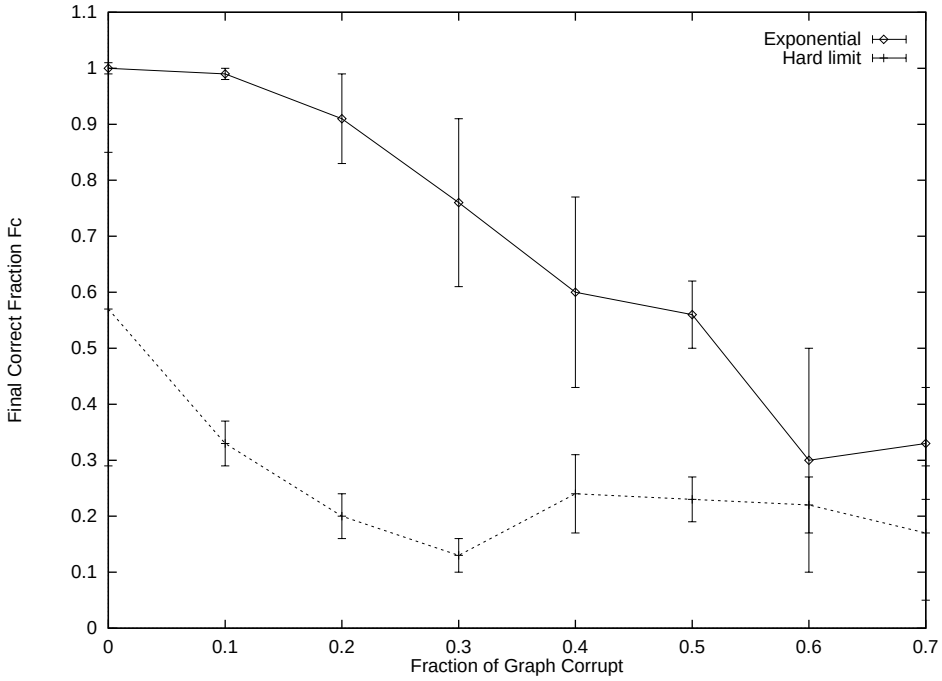


Figure 6.4: Performance of delta-function and exponential forms under corruption

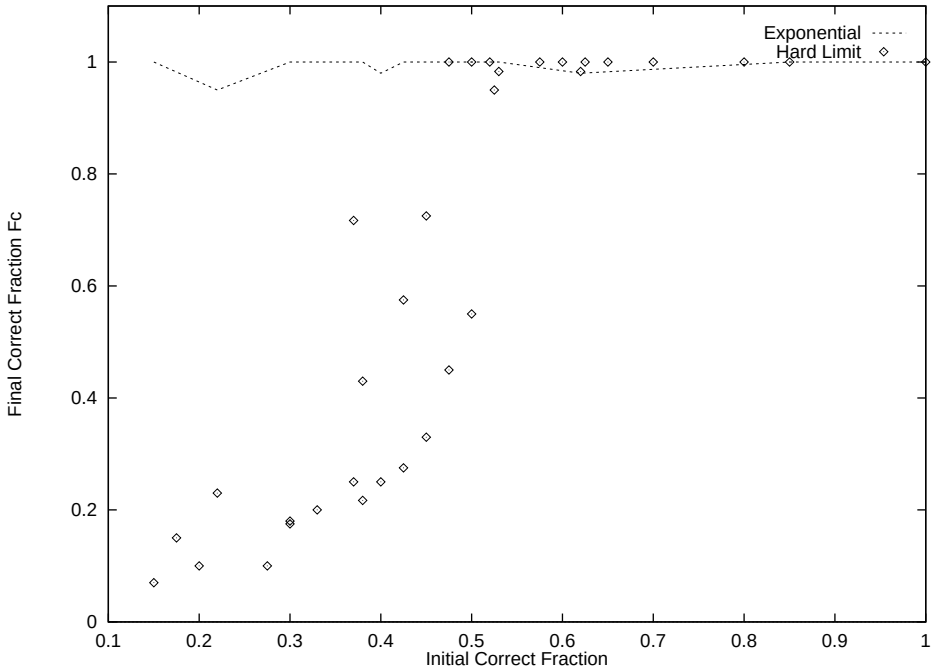


Figure 6.5: Relative performance of delta-function and exponential forms under initial errors

From this figure we can draw the conclusion that the delta-function limit operates well when the level of initialisation errors is low (in this experiment when the level of initialisation errors is less than 50%) and there are significant numbers of correctly matched relational units. At higher levels of initial errors, the scheme is unable to recover sufficient consistency to operate effectively. The full exponential scheme on the other hand is not sensitive to the level of initial errors.

6.5 Optimisation Schemes

There are a diversity of alternative optimisation schemes available in the literature. A particular energy function may often have features which require a specialised or complex update procedure to locate the optimal labelling configuration. In contrast we have chosen to update our functional by a simple gradient ascent procedure. In this section we compare and contrast this approach with some alternatives. In particular we wish to determine if there is a need for, or an advantage to be gained from, a more complicated approach.

6.5.1 Random Non-Deterministic Algorithm

This simple algorithm attempts to make a random label replacement at a random node within the graph. The replacement is accepted provided it increases the global consistency criterion. This effectively corresponds to a non-deterministic version of the gradient ascent method.

The aim of this experiment is to determining how susceptible the matching algorithm is to finding poor matches because of sub-optimal local maxima in the matching criterion. If there are a plethora of such maxima, a deterministic update procedure would be expected to ascend the gradient to one of these points, resulting in an inaccurate match. A random updating method would fall into a different maxima each time, resulting in varying matching performance over different runs of the algorithm on the same data.

Table 6.1 describes a set of results investigating a number of runs using both the deterministic and random algorithms on the same data. One set of data has been generated at each level of graph corruption. The deterministic algorithm is then run on this data to pro-

Corruption	Deterministic F_c	Deterministic $\mathcal{F}$	Random F_c	Random $\mathcal{F}$
0.0	1.0	0.0021	1.0	0.0021
			1.0	0.0021
			1.0	0.0021
0.1	0.87	0.0293	0.87	0.0293
			0.87	0.0293
			0.87	0.0293
0.2	0.73	0.056	0.73	0.056
			0.73	0.056
			0.73	0.056
0.3	0.6	0.00120	0.6	0.00120
			0.43	0.00037
			0.43	0.00037
0.4	0.2	0.000894	0.17	0.000887
			0.17	0.000892
			0.17	0.000892
0.5	0.13	0.000804	0.2	0.003040
			0.17	0.000805
			0.2	0.003040
0.6	0.2	0.00146	0.2	0.00146
			0.2	0.00146
			0.2	0.00146
0.7	0.125	0.000443	0.175	0.000439
			0.2	0.000438
			0.175	0.000439

Table 6.1: Comparison of Deterministic and Random Labelling

vide a benchmark against which the random approach can be compared, both in terms of the matching fraction achieved F_c, and the value of the criterion $\mathcal{F}$. The non-deterministic version is then run three times on the data to see whether a superior maximum of the criterion is found. These figures are shown in the fourth and fifth columns. From examination of the results, the random update approach shows superior performance on the graph with 50% corruption - the deterministic algorithm seems to have become stuck in a local maximum. On all the other runs the deterministic algorithm achieves the highest or equal highest value of the criterion.

It is interesting to note that in some cases the matching fraction is greater for the random algorithm where the criterion value is less than the deterministic algorithm. If the deterministic algorithm does find the global maximum of the criterion as suggested by these figures, then matching errors occur at higher levels of corruption because the optimum point does not correspond to the correct match; noise in the graph causes the maximum to be misplaced.

Of course this is rather an inefficient way to perform updates; in the experiments here the random algorithm was run with 3 to 5 times as many updates in total as the deterministic approach simply because the latter explores the possible label changes more efficiently.

6.5.2 Simulated Annealing

A more justifiable and interesting update procedure is provided by simulated annealing. This method which produces label updates in a non-deterministic way and allows label updates which reduce consistency is based on a Boltzmann machine interpretation of the labelling process. This analogy is developed more fully in section 5.3, where we show that the appropriate Gibbs potential is

$$U(\Gamma) = \frac{\sum_\pi \mathcal{H}(\Gamma,\pi) e^{-k\mathcal{H}(\Gamma,\pi)}}{\sum_\pi e^{-k\mathcal{H}(\Gamma,\pi)}} \tag{6.9}$$

Therefore in order to find the optimum labelling, we should anneal the labels in this potential field.

The annealing algorithm operates as follows. At each node u in G_1, a random match is chosen as a candidate update. The change in the Gibbs potential $U(\Gamma)$ is calculated for all cliques in which u resides; the change in energy $\Delta U = (1/|C_u|)\sum_{v\in C_u}[U(\Gamma'_v) - U(\Gamma_v)]$

is given by the average energy change over these cliques. The label update is accepted if $\Delta U \leq 0$. If $\Delta U > 0$ the update is accepted with a probability $e^{-k\Delta U}$. This procedure is repeated at each node in G_1. The whole sequence of node updates is then repeated $|V_1|$ times at a particular value of P_e to allow the labelling to approach an equilibrium state at a particular 'temperature'. This whole update sequence constitutes one iteration and evaluated the same number of possible label changes at the standard deterministic update algorithm, albeit in a random fashion.

The critical feature of this algorithm is it's ability to escape from local maxima and locate a globally optimal result, provided that a suitable temperature annealing schedule is adopted. According to Geman and Geman (Geman and Geman, 1984) the annealing 'temperature' T (corresponding to the parameter k in our scheme) should be reduced according to a theoretical schedule

$$T(i) = 1/k_e = \frac{N\Delta}{\log(1+i)} \tag{6.10}$$

where N is the number of sites of the system ($N = 30$), Δ is the difference between the maximum and minimum values of the energy function U ($\Delta \simeq 5.5$ for our criterion), and i is the iteration number. In the case of the graphs under study here, if for example we wish to reach $T = 0.2$ which is typical of the experiments presented earlier, then we would need e^{825} iterations. This schedule turns out to be prohibitive in practice (Geman and Geman, 1984). Rather, we reduce according to the schedule used previously, i.e.

$$P_e = P_e^{(0)} e^{-\alpha i} \tag{6.11}$$

The annealing schedule extends over 60 iterations, and α is chosen such that $T(60) = 0.2$ which corresponds to the same terminal value of k_e used in discrete relaxation experiments.

Figure 6.6 shows a comparison of the performance of the simulated annealing of the potential U with a gradient descent version also optimising U. This is identical to the simulated annealing algorithm except that updates which increase the value of U are not allowed. For comparison the exponential discrete method is also included.

These results are encouraging for the new energy function U; the performance of the gradient descent optimisation of U are very similar to that of the original discrete scheme.

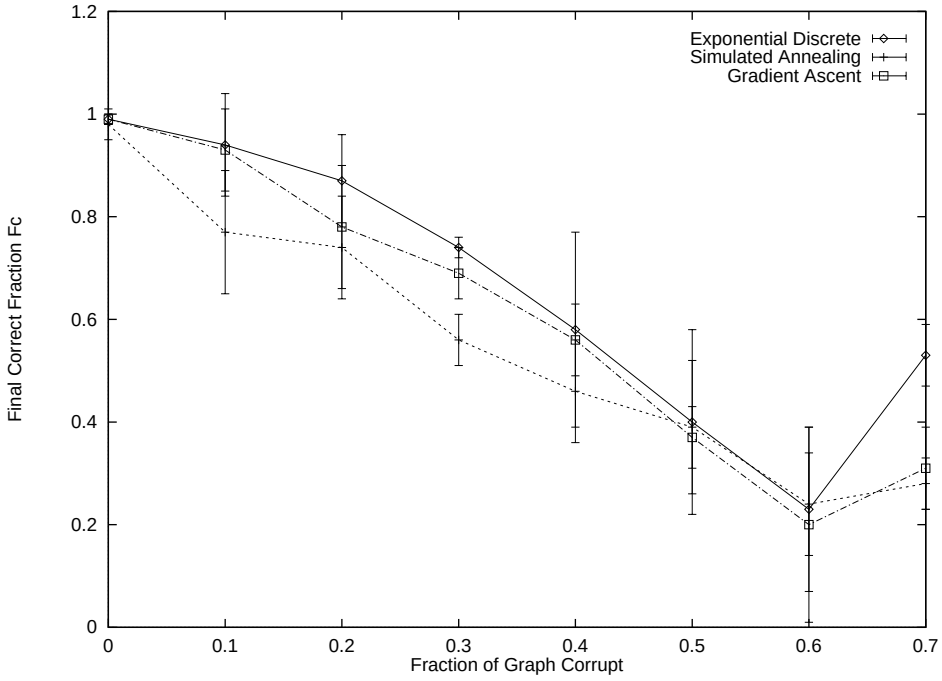


Figure 6.6: Relative performance of SA and GA on the Gibbs energy and exponential discrete relaxation

This supports the theoretical calculation of U as the corresponding energy representation of the global probability $P(f)$. The success of the gradient descent algorithm also suggests a lack of local maximum in U.

The simulated annealing approach is marginally worse than the other two algorithms, suggesting that even after 60 iterations convergence has not been reached. This is not surprising since in (Geman and Geman, 1984) 300-1000 iterations are suggested. However for the graph matching problem this number proves to be computationally prohibitive.

The conclusion from these results is that simulated annealing is unnecessary when optimising the energy function U since gradient ascent works satisfactorily, and that optimising U gives the same results as optimising the global probability $P(f)$.

6.6 Conclusions

A number of significant results come out of the studies in this chapter;

- Under a variety of limiting cases the compound exponential criterion can be approximated in terms of either a linear, mean square or delta function, which allows the

method to be related to other methods developed in the literature.

- Experimental studies show that all these approximations under-perform the full compound exponential criterion, but also reveal certain situations in which an approximate approach can be expected to be effective.

- In the second part of the chapter we compare a random update algorithm and simulated annealing with the gradient ascent optimisation approach. The conclusion of this study is that gradient ascent optimises the consistency as effectively as either of the other two methods and does not become stuck in local minima.

Chapter 7

Probabilistic Relaxation

In the previous two chapters we have concentrated on finding the MAP estimate of the joint node mapping probability. Essentially the MAP labelling scheme takes a global view of the match, often referred to as a message centred approach. In this philosophy we considered the matching of all objects at the same time while using only local measurement information. This model leads to a matching criterion on which a discrete MAP optimisation scheme operated.

In this chapter we will focus on another modality of decision making, the object centred approach. In this we confine ourselves to matching a single object at a time, using the measurement information available for all objects. Following this model leads us to a probabilistic relaxation scheme which combines evidence for a match from the neighbourhood to compute continuous matching probabilities. Using a similar modelling philosophy for matching errors to that previously described in the construction of our consistency criterion, we show that such a framework allows the calculation of compatibility coefficients which are devoid of parameters. Finally we examine the comparative effectiveness of the new probabilistic relaxation scheme and the discrete relaxation approach.

7.1 Introduction

Probabilistic relaxation was originally conceived by Rosenfeld, Hummel and Zucker (Rosenfeld et al., 1976) as a continuous labelling extension to the work of Waltz on discrete relaxation

labelling (Waltz, 1975). As such it is concerned with combining evidence from neighbouring objects, and has traditionally been viewed as a low-level vision process. Since that seminal paper, a plethora of alternatives have evolved. Variants of the original probabilistic scheme (Lloyd, 1983; Ton and Jain, 1989; Tang and Lee, 1992; Chipman and Ranganath, 1992) have dominated the literature. This growth of alternative formulations was driven by the fact that in Rosenfeld, Hummel and Zucker's original scheme there was no theoretical grounding to the definition of a support function, leaving the path clear for an explosion of essentially heuristic support formulations. Kittler and Hancock (Kittler and Hancock, 1989) have recently overcome some of the problems of Rosenfeld, Hummel and Zucker's original scheme by casting the process in a rigorous Bayesian framework. One important result of this framework is an exact expression for the support function in terms of label configuration probabilities. From this basis support functions for specific problems can be derived in a principled way by defining the appropriate probability distributions. Building on this work, Kittler, Christmas and Petrou (Kittler et al., 1993) have shown how attributed relational graphs may be matched by extending the probabilistic framework to incorporate binary measurements between pairs of nodes.

Another feature of the Kittler-Hancock support function in its original form is the complexity; the evaluation of support involves the calculation of an exponential number of terms. However Kittler and Hancock (Kittler and Hancock, 1989) present two methods of simplifying the computational complexity of the support function. The first method involves the specification of a dictionary of allowed labellings, with configurations only considered if they reside in the dictionary. The support then need only be evaluated over this dictionary of labellings, reducing the complexity to the size of the dictionary. The second method consists of making independence assumptions between the nodes; with suitable node inter-dependencies the support function may be factorised and simplified. This approach in effect reduces the size of contextual units to pairs or triplets of nodes.

For reasons of efficiency and robustness, and their intrinsically parallel nature, relaxation schemes make an attractive choice for relational matching problems. However when considered in the light of relational matching tasks, the algorithms reported in the literature suffer a number of deficiencies. The relaxation methodology originated in the labelling of low-level

image entities such as pixels and, despite advances, the basic method remains orientated towards early vision. The method is low-level in terms of both concept and modelling ingredients, and as such it lacks suitability and representational power at the intermediate and high levels. In the past relaxation has been applied mainly to pixel based tasks such as image restoration(Geman and Geman, 1984; Milun and Sher, 1993), and segmentation(Cohen and Cooper, 1987; Kittler and Hancock, 1989; Hancock et al., 1992). Schemes also have a tendency to be difficult to control with a number of parameters which must be manually adjusted and have no obvious meaning in terms of the scene under match. Above all none model the corruption processes at work in the formation of relational graphs from noisy and uncertain data.

In this chapter we will apply a similar framework for modelling node corruption used previously in Chapter 2.15 on a restricted range of graph configurations. This restriction on node interdependencies will allow us to simplify the problem somewhat by factorising the support function.

7.2 Decision making in Probabilistic relaxation

The probabilistic relaxation method is derived from an object centred interpretation and as such attention is confined to finding the optimal label assignment for a single element at a time using the entire set of measurement information. From the perspective of information theory, the objective is to maximise the *a posteriori* probability $P(f(u)|\mathbf{X})$. The appropriate Bayesian decision rule is therefore

$$f(u) = \arg\max_{w \in V_2} P(f(u) = w|\mathbf{X}) \tag{7.1}$$

This contrasts with our previous approach; before we were interested mainly in the joint prior $P(f)$ and the single label aposteriori probability $P(f(u)|\mathbf{x}_u)$. Information was provided by the mapping function and from the graph structure, with measurement information entering only in a local sense through the single label probabilities. The probabilistic relaxation approach is concerned instead with combining measurement information in a more extensive way and as such would seem to be at odds with the discrete interpreta-

tion. However, it is still graph structure which determines how information propagates between nodes. This being the case, structural corruption models are still important in the construction of a probabilistic relaxation scheme. As we shall discover, there are some interesting results to be obtained concerning rôle of graph structure in limiting the feasibility of matching by this method.

At this stage we assume that only local interactions are significant. Only nodes connected to each other by a graph arc will interact. This simplifies the conditional probability of interest to

$$P(f(u)|\mathbf{X}) = P(f(u)|\mathbf{x}_v \forall v \in C_u) \tag{7.2}$$

7.3 The Support Function

The probability of Equation 7.1 provides the starting point from which Kittler and Hancock (Kittler and Hancock, 1989) have developed an evidence-combining formula reminiscent of the classic Rosenfeld, Hummel and Zucker (Rosenfeld et al., 1976) probabilistic relaxation formula.

$$P^{(n+1)}(f(u)) = \frac{P^{(n)}(f(u))Q^{(n)}(f(u))}{\sum_v P^{(n)}(f(v))Q^{(n)}(f(v))} \tag{7.3}$$

The development of the relaxation formula is based on the concept of iteratively filtering measurements to reinforce the consistency of the labelling. The filtered measurements are not explicitly calculated but they are implicitly specified in conditional probabilities. This approach leads to a probability update rule (Equation 7.3) with the same structure as that of Rosenfeld Hummel and Zucker.

The crucial ingredient in this update formula is the support function $Q(f(u))$ which combines evidence from the neighbourhood $C_u^{(1)}$ of node u for the match $u \rightarrow f(u)$. In our graph formulation this neighbourhood consists of the set of nodes connected to u by an arc. These nodes interact directly with node u via the graph topology - they share a common arc. The support function Q is specified in many different ways in the literature, commonly they are heuristic (Izumi et al., 1992; Ton and Jain, 1989; Tang and Lee, 1992), have internal

inconsistencies (Chipman and Ranganath, 1992; Ton and Jain, 1989) or are only valid in the limit of weak contextual information (Kittler et al., 1993; Ton and Jain, 1989; Ziqing Li, 1990). We advocate the use of the internally consistent support function of Hancock and Kittler (Kittler and Hancock, 1989). Details of the derivation are beyond the scope of this thesis, instead we quote the expression produced by Kittler and Hancock:

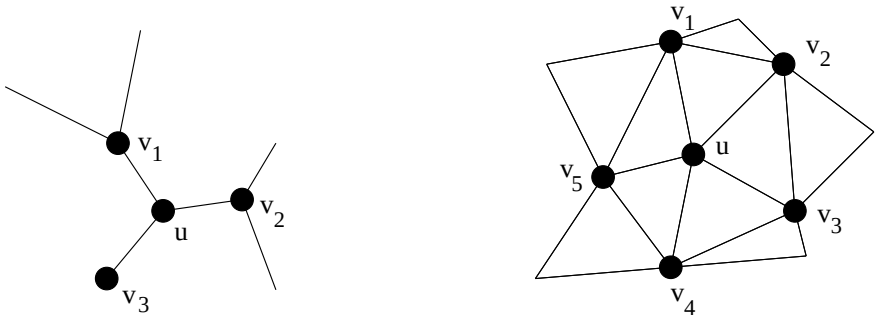
$$Q^{(n)}(f(u)) = \frac{1}{P(f(u))} \sum_{f(C_u^*)\in\Omega} \Big\{ \prod_{v\in C_u^*} \frac{P^{(n)}(f(v))}{P(f(v))} \Big\} P(f(w)\forall w \in C_u^{(1)}) \tag{7.4}$$

where Ω is the possible label configuration space for C_u^*, i.e. all possible labellings on all nodes of C_u^*, and C_u^* represents clique $C_u^{(1)}$ excluding the centre node u. It is on this support function that we will now focus our attention. We wish to build a model based on the same principles on which we founded our clique probabilities $P(\Gamma)$. From this model we will derive the compatibility coefficients appropriate to various graph matching tasks.

There are a number of important elements to this expression; the reader should note that the sum is over the entire configuration space of C_u^*. The support function is therefore of the same order of complexity as the graph matching problem for $C_u^{(1)}$ (i.e. the sum has an exponential number of terms). The support function comprises of the prior probabilities of matches $P(f(u))$ reflecting world knowledge of the chance of individual matches;the probability $P^{(n)}(f(u))$ reflecting the current state of the match; and the joint prior $P(f(w)\forall w \in C_u^{(1)})$ which represents the world model of context. It is this last term which provides the structural information to the scheme and is the key modelling element.

7.4 Factorisation

The support function above (Equation 7.4) has exponential complexity and is not suitable for use in realistic matching tasks. In order to simplify the expression sufficiently a number of techniques are available, including the weak context approximation, dictionary-based techniques or factorisation. We will apply the factorisation ideas of Kittler and Hancock to avoid having to enumerate the configurations in the dictionary. This approach involves the factorisation of the joint prior $P(f(v)\forall v \in C_u^{(1)})$ by expanding in terms of conditional label probabilities. Under certain limiting assumptions these conditional probabilities can



a) Tree-like: Pairwise interactions b) Triangular: Triplet interactions

Figure 7.1: Example Graph Types: a tree-like graph and a triangulation

be simplified to obtain a result in terms of interactions of a lower order than the size of the clique, for example node-pairs or triplets. However the exact factorisation scheme we should use depends entirely on the interactions between elements in neighbourhood $C_u^{(1)}$ and hence on the structure of the graph itself. In order to select the appropriate scheme we must first limit the discussion to a specific graph structure. Here we develop the simplest case, that of a tree-like graph. In this structure there are no direct interactions between the external nodes of a clique (Figure 7.1a). The reader should note that in terms of the experimental data this corresponds to the structure of the road network graphs.

If we label the nodes making up the clique $C_u^{(1)}$ as $\{u, v_1, v_2, \ldots, v_N\}$ where $N = |C_u^{(1)}|$, then by appliying the Bayes rule, the factorisation proceeds as follows:

$$\begin{aligned} P[f(u), f(v_1), \ldots, f(v_N)] &= P[f(v_1)|f(u), f(v_2), \ldots, f(v_N)] \\ &\quad \times P[f(v_2)|f(u), f(v_3), \ldots, f(v_N)] \times \ldots \\ &\quad \times P[f(u)] \end{aligned} \tag{7.5}$$

where u is the central node. Since the external nodes are independent of each other, the conditional probabilities simplify thus: $P[f(v_1)|f(u), f(v_2), \ldots, f(v_N)] = P[f(v_1)|f(u)]$. The final expression for the joint probability is

$$P[f(u), f(v_1), \ldots, f(v_N)] = P[f(u)] \prod_{n=1}^{N} P[f(v_n)|f(u)] \tag{7.6}$$

Substitution of Equation 7.6 into the support function in Equation 7.4 gives

$$Q^{(n)}(f(u)) = \frac{1}{P(f(u))} \sum_{f(C_u^*)\in\Omega} \Big\{ \prod_{v\in C_u^*} \frac{P^{(n)}(f(v))}{P(f(v))} \Big\} P[f(u)] \prod_{n=1}^{N} P[f(v_n)|f(u)] \tag{7.7}$$

Since the sum covers all labellings on C_u^*, we may factorise the sum and interchange the sum and product to obtain

$$Q(f(u)) = \prod_{m=1}^{N} \sum_{f(v_m)\in V_2} \frac{P(f(u)|f(v_m))}{P(f(u))} P^{(n)}(f(u)) \tag{7.8}$$

In other words the support function is considerable simplified into a product over the node neighbourhood $C_u^{(1)}$ a form more amenable to computation.

7.5 Compatibility Coefficients

In this section we calculate expressions for the various probabilities appearing in Equation 7.8 using the modelling philosophy in Chapter 2. In order to clarify the rôle of the various terms in Equation 7.8, we begin by re-writing the support function in terms of a compatibility coefficient R and the unary matching probability $P^{(n)}$:

$$Q(u \to v) = \prod_{m=1}^{N} \sum_{v_m\in V_2} P^{(n)}(f(v_m)) R(f(u), f(v_m)) \tag{7.9}$$

The compatibility coefficient R is specified by the mutual information measure

$$R(f(u), f(v_m)) = \frac{P(f(u), f(v_m))}{P(f(u))P(f(v_m))} \tag{7.10}$$

It is this measure of contextual information which is fundamental to the calculation of support. In simple terms it represents the ratio of the probability of the two matches appearing on adjacent nodes to the apriori probability of the matches appearing in isolation. It is this ratio which provides the vehicle for contextual information to influence the matching process. The contextual information is based purely on binary relationships between the pairs of matches $f(u)$ and $f(v_m)$ on the edges in the graph. If nodes u and v_m are linked by an arc in G_1 then we would correspondingly expect their matches $f(u)$ and $f(v_m)$ to also be linked by an arc in G_2, and so the probability of that particular pair of matches is enhanced with respect to the apriori probabilities. Similarly two neighbouring nodes whose matches

are not neighbours in G_2 are less likely to occur. In order to complete the relaxation scheme we need to build a quantitative model of the probabilities of these various occurrences.

We begin by noting that graph corruption will disrupt this ideal pattern by introducing spurious nodes and arcs. In order to accommodate the possibility that objects in the graph are produced by noise and segmentation error, we must allow for the presence of unmatchable nodes. We therefore augment the graphs with a null node ϕ to which unmatchable entities may be assigned. Furthermore we assume that node corruption errors occur with a uniform and memoryless probability p across the entire graph.

It is important at this point to stress some of the limitations of this model. We have assumed that the only process at work in corrupting the topology of the graph is that of the introduction of extraneous nodes. The relations themselves considered to be uncorrupted; if the nodes in a portion of the graph are undisrupted, the set of edges will be identical. This is not valid in the case where relations between objects are difficult to extract. We have also not considered explicitly the case of node dropout, but if the condition of uncorrupted relations is upheld, the loss of nodes will not affect the matching of the remaining structure. The edges between the uncorrupted nodes will remain undisrupted.

The corruption may also not be uniform across the scene - factors such as distortion and noise may cause a particular region to be particularly poorly segmented. The approximation that we have adopted can be expected to perform poorly on localised patches of increased corruption.

7.5.1 Scene to Model Matching

In this subsection we will concern ourselves with matching a scene graph to a model graph. The model graph is assumed to be uncorrupted and therefore we do not need to consider the effect of any spurious or missing nodes or arcs in it's structure.

Uncertainty in the graph representation is present in two separate forms. Geometric distortion causes variation in measurement information such as angle and length. Topological information about the connectivity of the relational graph may be left unchanged by such distortion, but this is largely dependent on the type of relation. As mentioned above the adopted model does not account for purely relational corruption. For example corner

relations will be significantly affected by geometric distortions in the image, whereas the Delaunay graph is robust to such distortions (Finch et al., 1995; Tuceryan and Chorzempa, 1991). If matching is to be accomplished by structural means it is critical to employ such robust representations. Poor node segmentation will still corrupt the relational description. Following our original philosophy of relying on structural information to provide context, we propose a dichotomy between the two types of information - geometric information is used exclusively in the initial probability model, and topological information only is used in the modelling of constraints during the relaxation process. Since the topology of the graphs is represented purely by the interconnectivity of nodes and arcs, this model is purely symbolic.

The topology of the graphs are used to provide constraints in the following manner; if two nodes are linked by an arc in graph G_1 then there are two possible situations. If the graph G_1 is uncorrupted in that region the arc will have a matching arc in the model graph. If, on the other hand, either of the nodes correspond to a segmentation error then the arc relation between them will have no corresponding match and one or both of the nodes is unmatchable. These unmatchable nodes must be labelled as null (ϕ). If segmentation errors occur with a uniform memoryless probability p then a probability mass $(1-p)^2$ is associated with uncorrupted node pairs, $p(1-p)$ with one corruption and p^2 if both nodes correspond to corruption (and must consequently be labelled by the label ϕ as unmatchable). If neither node has been labelled as corruption and the corresponding model arc does not exist, the configuration is inconsistent and is disallowed entirely (Figure 7.2).

We can then expand the node-pair probability over the appropriate set of constraints. For example if a consistent arc is being considered, the set of edges in G_2 is the relevant set of constraints, i.e. $P[f(u), f(v)] = \sum_{E \in E_2} P[(u,v)|E]P(E)$. The conditional probability $P[(u,v)|E]$ is specified by the distribution rule explained above. For the prior $P(E)$ we adopt a simple uniform distribution assuming no apriori knowledge of the edge: $P(E) = \frac{1}{|E_2|}$.

If u and v_m are linked by an arc, the following probabilities apply

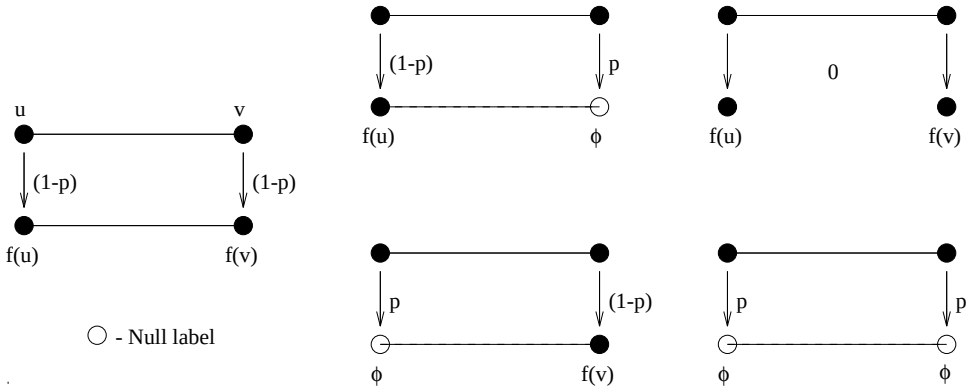


Figure 7.2: Possible mappings between the nodes forming two edges

$$P(f(u), f(v_m)) = \begin{cases} \frac{(1-p)^2}{|E_2|} & \text{if } (f(u), f(v_m)) \in E_2 \\ \frac{p(1-p)}{|V_2|} & \text{if } (f(u), f(v_m)) \in (V_2 \times \phi) \cup (\phi \times V_2) \\ p^2 & \text{if } (f(u), f(v_m)) \in (\phi \times \phi) \\ 0 & \text{if } (f(u), f(v_m)) \in (V_2 - \phi) \times (V_2 - \phi) - E_2 \end{cases} \tag{7.11}$$

The axioms of probability require that $\sum_{\forall m,n} P(f(u), f(v_m) = w_n) = P(f(u))$, and the single label priors required in Eqn. 7.10 are calculated in this fashion to be

$$P(f(u)) = \begin{cases} \frac{1-p}{|V_2|} & \text{if } f(u) \in V_2 \\ p & \text{if } f(u) = \phi \end{cases} \tag{7.12}$$

Substituting equations 7.12 and 7.11 into equation 7.10, the compatibility coefficients are specified by

$$R(f(u), f(v_m)) = \begin{cases} \frac{|V_2|^2}{|E_2|} & \text{if } \{f(u), f(v_m)\} \in E_2 \\ 1 & \text{if } \{f(u), f(v_m)\} \in (\phi \times V_2) \cup (V_2 \times \phi) \cup (\phi \times \phi) \\ 0 & \text{if } \{f(u), f(v_m)\} \in (V_2 - \phi) \times (V_2 - \phi) - E_2 \end{cases} \tag{7.13}$$

The probability of node corruption cancels from the compatibility coefficients; the entire graph-based constraint process is captured by a model which is entirely devoid of free parameters; it is specified purely in terms of the numbers of arcs and nodes in the model graph. This is an important result; it allows the matching of corrupt graphs without the need for knowledge of either the source or level of node corruption. The strength of the constraints is based purely on the structure of the graphs under match. Infact the compatibility of a

consistent edge is given by the edge density of the graph. The pattern of compatibilities also has a pleasing intuitive feel; the power of an edge constraint increases as the number of edges decreases. If a null label is present and no edge information exists, the information provided is neutral (the compatibility is unity), and any inconsistent arc is completely disallowed (compatibility zero).

7.5.2 Scene to Scene Matching

In the case of matching scene graph to scene graph, we must additionally account for possible corruption of graph G_2. In this case V_1 is also augmented with a null label, and compatibility is computed taking into account the possible corruption of any of nodes involved - uncorrupted items appear with a probability $(1-p)^4$. Furthermore the relevant constraint class is drawn from both graphs i.e. $E_1 \times E_2$. Following the methodology in section 7.5.1,

$$R(f(u), f(v_m)) = \begin{cases} \frac{|V_1 \times V_2|^2}{|E_1 \times E_2|} & \text{if } ((u, v_m), (f(u), f(v_m))) \in E_1 \times E_2 \\ \frac{|V_1|^2}{|E_1|} & \text{if } ((u, v_m), (f(u), f(v_m))) \in E_1 \times \Phi_2 \\ \frac{|V_2|^2}{|E_2|} & \text{if } ((u, v_m), (f(u), f(v_m))) \in \Phi_1 \times E_2 \\ 1 & \text{if } ((u, v_m), (f(u), f(v_m))) \in \Phi_1 \times \Phi_2 \\ 0 & \text{otherwise} \end{cases} \tag{7.14}$$

where $\Phi_n = \{(V_n \times \phi) \cup (\phi \times V_n) \cup (\phi \times \phi)\}$ is the set of label pairs containing a null label. The compatibility coefficients are now symmetrical with respect to the swapping of the scene graphs as intuitively they must be. However the normalisation of probabilities implied in Eqn. 7.3 is not symmetrical since the method was originally conceived as an object labelling problem. However in the limit of a hard consistent labelling, the symmetry condition on the probability normalisation also holds.

At first glance it seems curious that by admitting the possibility of node corruption in G_2 at any level, the consistent edge compatibility is increased when compared to the original scene to model compatibilities, to $\frac{|V_1 \times V_2|^2}{|E_1 \times E_2|}$. In the limit of small corruption probability we would expect the values of compatibility to be identical. However the two cases examined in Section 7.5.1 of a consistent edge (compatibility $\frac{|V_2|^2}{|E_2|}$) and null labels (compatibility 1) should be compared to the constraint sets $E_1 \times E_2$ and $E_1 \times \Phi_2$ respectively. It can then be

seen that the ratio of compatibilities remains unchanged.

7.6 Matching Delaunay Triangulations

Finally we briefly study the matching of triangulated Delaunay graphs using this probabilistic framework. The work reported here is based on the paper by Finch, Wilson and Hancock, "Matching Deformed Delaunay Triangulations" (Finch et al., 1995) and is of interest here for two reasons. Firstly, it draws on the same error models discussed above. Secondly, the work provides a prime example of how the factorisation scheme must be tailored to the specific type of graph structure in use. Moreover, it emphasises the increased complexity created by the use of interacting triplets rather than pairs.

While the use of pairwise node interactions is valid for tree-like graphs, application of the same scheme to Delaunay triangulations not only ignores a wealth of addition constraints and limits performance, but also involves an inconsistent treatment of node interactions. In effect we are ignoring the local topology of the Delaunay graph; such a graph consists entirely of triangular faces and mutual interactions of the external nodes exist. Furthermore, use of a pairwise scheme can lead to a matched graph which, while it contains consistently matched edges, contains incorrectly ordered graph neighbours and is not itself triangulated. In order to overcome these difficulties we must re-examine the assumptions which allowed us to simplify the original factorisation of the joint probability in equation 7.6.

Reference to section 7.4 shows that previously we assumed a simplification of the conditional probabilities based on independence of the external nodes in the neighbourhood (i.e. non-centre nodes). Here the external nodes are actually inter-dependent, and this model no longer applies. Instead the following model is necessary:

$$P(f(v)|f(w), w \in C_u^{(1)}, w \neq v) = P(f(v)|f(w), w \in C_u^{(1)}, w \in C_v^{(1)}, w \neq v) \tag{7.15}$$

In other words, two external nodes v and w interact with each other provided v is in the neighbourhood of w. In order to make use of this simplifying assumption we must re-factorise the joint prior in a different way:

$$
\begin{aligned}
P(f(u), f(v_1), \ldots, f(v_N)) &= P(f(v_N)|f(u), \ldots, f(v_{N-1})) \\
&\quad P(f(v_{N-1})|f(u), \ldots, f(v_{N-2})) \ldots \\
&\quad P(f(v_1)|f(u))P(f(u))
\end{aligned} \tag{7.16}
$$

Using Bayes rule and Equation 7.15 this expression may be rewritten in terms of interacting triplets. See (Finch et al., 1995) for a mathematically detailed treatment. The resulting expression is

$$
\begin{aligned}
P(f(u), f(v_1), \ldots, f(v_N)) &= P(f(v_1)|f(u), f(v_N)) \\
&\quad P(f(v_2)|f(v_1), f(u)) \ldots \\
&\quad P(f(v_N)|f(v_{N-1}), f(u))
\end{aligned} \tag{7.17}
$$

and thus the joint probability appearing in the support function can be expressed in terms conditional probabilities involving triples of node matches on graph triangles. However this triplet pattern has an inter-dependence which prevents the restructuring of the support function into a product over the nodes in the neighbourhood of u, and adds considerably to the complexity of the expression. Expanding the support function over the set of labellings Ω, we find

$$
\begin{aligned}
Q(f(u)) = \frac{1}{P(f(u))} \quad & \sum_{f(v_1)\in V_2} \frac{P^{(i)}(f(v_1))}{P(f(v_1))} \sum_{f(v_n)\in V_2} P(f(v_1)|f(v_n), f(u)) \\
& \sum_{f(v_2)\in V_2} \frac{P^{(i)}(f(v_2))}{P(f(v_2))} P(f(v_2)|f(v_1), f(u)) \ldots \\
& \sum_{f(v_k)\in V_2} \frac{P^{(i)}(f(v_k))}{P(f(v_k))} P(f(v_k|f(v_{k-1}), f(u)) \\
& \sum_{f(v_{k+1})\in V_2} \frac{P^{(i)}(f(v_{k+1}))}{P(f(v_{k+1}))} P(f(v_{k+1})|f(v_k), f(u)) \ldots \\
& \sum_{f(v_n-1)\in V_2} \frac{P^{(i)}(f(v_n-1))}{P(f(v_n-1))} P(f(v_n-1)|f(v_n-2), f(u)) \\
& \frac{P^{(i)}(f(v_n))}{P(f(v_n))} P(f(v_n)|f(v_n-1), f(u))
\end{aligned} \tag{7.18}
$$

Details of how to evaluate this support function in a computationally efficient way are not our primary concern here; they are discussed more fully in (Finch et al., 1995). The primary interest here is the development of a suitable compatibility function. The function is still of exponential complexity but Equation 7.18 may be implemented in a recursive fashion to reduce the complexity to $O(nm^3)$. This implementation corresponds to evaluating the labelling of graph triangles in a cyclic fashion around the central node.

7.6.1 Triplet Compatibility Coefficients

Again we re-write the support function in terms of a compatibility coefficient R specified by the mutual information measure thus:

$$\begin{aligned} R(f(v_k), f(v_{k-1}), f(u)) &= \frac{P(f(v_k|f(v_{k-1}), f(u))}{P(f(v_k))} \\ &= \frac{P(f(v_k), f(v_{k-1}), f(u))}{P(f(v_k))P(f(v_{k-1}), f(u))} \end{aligned} \tag{7.19}$$

The reader should note the similarities between this expression and the previous edge-based expression (equation 7.10). This time we are interested in the relative probability of a graph triangle to that of a radial edge and an isolated node. We now wish to model these triplet compatibility coefficients following the model described in section 7.5. Following this methodology, a binomial distribution of node corruption probabilities is adopted. In the triplet scenario the configurations are as follows;

- all nodes from the face in V_1 are matched to a valid face in the model graph. This occurs with probability $(1-p)^3$.
- Two nodes from the face match to an edge in V_2 and the other node is null-matched. This occurs with probability $p(1-p)^2$.
- Two nodes are null matched and the third matches to any node in V_2. This occurs with probability $p^2(1-p)$.
- All three nodes are null-matched. In this case the probability is p^3.

All other configurations are forbidden and account for zero probability mass. Again the probability mass is uniformly distributed over the relevant class of configurations. Consequently the joint triplet probabilities obey the following distribution rule

$$P(f(u),f(v),f(w)) = \begin{cases} \frac{(1-p)^3}{|F_2|} & \text{if } \{f(u),f(v),f(w)\} \in F_2 \\ \frac{p(1-p)^2}{|E_2|} & \text{if } \{f(u),f(v)\} \in E_2 \text{ and } f(w) = \phi \\ \frac{p^2(1-p)}{|V_2|} & \text{if } f(u) \in V_2 \text{ and } f(v) = \phi \text{ and } f(w) = \phi \\ p^3 & \text{if } f(u) = \phi \text{ and } f(v) = \phi \text{ and } f(w) = \phi \\ 0 & \text{otherwise} \end{cases} \tag{7.20}$$

where F_2 is the set of triangular faces in G_2.

The pairwise distribution rule is identical to that specified previously in Equation 7.11. Again the single-label priors in the denominator of equation 7.19 can be obtained in the axiomatic way by summing the joint probabilities and are identical to the previously explored case (Equation 7.12). The compatibility coefficients are found by substituting equations 7.11, 7.12 and 7.20 into equation 7.19. They are given by the following rule

$$R(f(u),f(v_{k-1}),f(v_k)) = \begin{cases} \frac{|E_2||V_2|}{|F_2|} & \text{if } \{f(v_k),f(v_{k-1}),f(u)\} \in F_2 \\ \frac{|V_2|^2}{|E_2|} & \text{if } \{f(v_{k-1}),f(u)\} \in E_2 \text{ and } f(v_k) = \phi \\ & \text{or if } \{f(v_{k-1}),v_k\} \in E_2 \text{ and } f(u) = \phi \\ 1 & \text{if } f(u) \in V_2 \text{ and } f(v_{k-1}) = f(v_k) = \phi \\ & \text{or if } f(v_{k-1}) \in V_2 \text{ and } f(u) = f(v_k) = \phi \\ & \text{or if } f(v_k) \in V_2 \text{ and } f(u) = f(v_{k-1}) = \phi \\ & \text{or if } \{f(v_k),f(u)\} \in E_2 \text{ and } f(v_{k-1}) = \phi \\ & \text{or if } f(u) = f(v_{k-1}) = f(v_k) = \phi \\ 0 & \text{otherwise} \end{cases} \tag{7.21}$$

Again the constraint process and compatibility coefficients are captured without the need for parameters. The coefficients are based entirely on the global topological properties of the graphs. These coefficients and the matching configurations they correspond to are displayed in Figure 7.3. This pattern of compatibilities has a number of features worthy of comment. Firstly the compatibility coefficients grade the different face constraints according to their overall consistency. For example, in a fully triangulated planar graph, a consistently

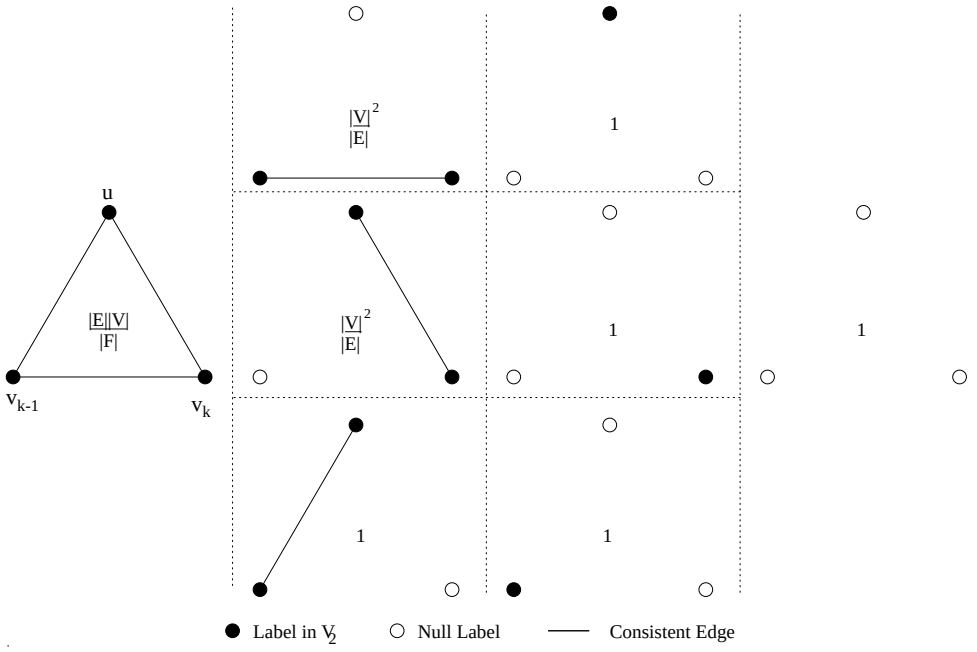


Figure 7.3: Allowed face configurations in a triangulated graph

matched face has a larger compatibility value than a consistently matched edge. The pattern of compatibilities associated with isolated edges is more complicated, but can be viewed as discouraging violation of the ordering constraint on external nodes of the clique - in a planar graph only cyclic permutations of the external nodes are valid mappings (Wang and Abe, 1995). This is most clearly evident when the consistent edge is between external nodes (case 3 in Figure 7.3). The compatibility in this case is large, encouraging external edges which enforce external node ordering. As for the other edge compatibilities, a trailing consistent radial edge is favoured over a leading one; this pattern discourages the introduction of a new edge which is potentially mis-ordered with respect to the node v_{k+1} in the following triplet in favour of a trailing edge whose ordering has already been evaluated with respect to node v_{k-2}. It is these ordering constraints which enhance the power of the triplet scheme over the original edge-based relaxation scheme.

7.7 Relevance of Graph Structure to Constraints

One of the more interesting features of the compatibility models presented above is their dependence solely on graph structure. We can use this fact to elucidate the relationship

between graph structure and the power of constraints provided. The compatibility coefficient gives us a direct measure of the power of the constraints provided to the probabilistic relaxation algorithm. To explore this relationship, we will consider a number of special graph structures and evaluate their corresponding compatibilities.

7.7.1 A Fully Connected Graph

In this graph structure all nodes are connected to all others. There are $|V|(|V|-1)/2$ edges in such a graph, and $|V|(|V|-1)(|V|-2)/3$ faces. The compatibilities are therefore given by

$$R_E = \frac{|V|^2}{|V|(|V|-1)/2} \simeq 2 \tag{7.22}$$

for the consistent edges, and

$$R_F = \frac{|V|^2(|V|-1)/2}{|V|(|V|-1)(|V|-2)/3} \simeq \frac{3}{2} \tag{7.23}$$

for the consistent faces. In this situation R takes on it's lowest value and constraints are at their weakest. However even in this regime the compatibilities are always too large to justify a weak-context approximation which is only valid for $R \simeq 1$. Finally it is interesting to note that the face compatibility is lower than the edge; the triplet model is little use on this type of graph. However we should be wary about applying the face model to such a graph since the factorisation and recursive evaluation of the triplet support function is dependent on the graph being planar.

7.7.2 The Delaunay Graph

The Delaunay graph is a neighbourhood graph derived from the Voronoi tessellation (Ahuja et al., 1985) of the image plane. Empirically, each node has 5.5 neighbours on average, and the number of graph edges is therefore $2.75|V|$. Since the image plane is fully triangulated, each edge participates in two faces, with each face requiring three bounded edges. Corresponding there are $5.5|V|/3$ faces in the graph. Substitution in Equation 7.21 gives

$$R_E = \frac{|V|^2}{2.75|V|} \simeq 0.36|V| \tag{7.24}$$

for the consistent edges, and

$$R_F = \frac{2.75|V|^2}{5.5|V|/3} \simeq 1.5|V| \tag{7.25}$$

for the consistent faces. Clearly with this relational model, not only are the compatibilities much larger than for the fully connected case, but the face model is also the dominant influence.

7.7.3 The Tree-like Graph

In this graph type, the nodes are connected in a tree structure. This corresponds to the minimally connected non-disjoint graph. There are $|V| - 1$ graph edges and no graph triangles. The face model is not appropriate for this type of graph, and we must turn our attention to the pairwise model. The edge compatibility is given by

$$R_E = \frac{|V|^2}{|V| - 1} \simeq |V| \tag{7.26}$$

It is in this situation that the edge constraints are at their strongest. This graph structure is exactly that produced by the road network data and therefore the edge-based scheme is the appropriate model for that data.

7.8 Evaluation of Probabilistic Relaxation

In order to evaluate the effectiveness of this scheme we present a number of experiments with the same datasets as described in Chapter 3. As before we quantify matching performance against a hand-generated ideal by specifying the fraction of the correct matches. In addition we enumerate a new measure of the ability of the algorithm to reject incorrect matches as noise. This is defined as

$$F_n = \text{incorrect matches}/\text{non} - \text{null matches}$$

The quantity $F_n \in [0, 1]$ gauges the amount of noise in the match, 0 being the best performance and 1 being the worst.

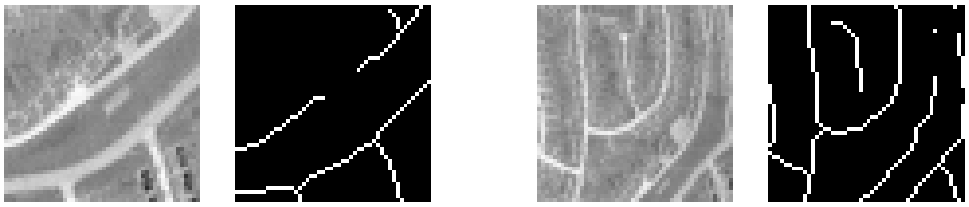


Figure 7.4: Examples of the appearance of spurious line-endings

7.8.1 Road Networks

The road network data used in the first study is presented in Appendix A. This data is ideally suited to our needs in that it closely fulfils the conditions set out in section 7.4; the road network is tree-like by design in order to join the nodes with the lowest edge density possible. The graph has the same topological structure as the road network itself and is essentially tree-like in nature. As discussed earlier, a tree-like graph provides the most powerful set of constraints for the edge-based algorithm to work with. Furthermore spurious road connections are rarely detected. However the corruption processes which affect the graph generate a plethora of spurious line-endings, while non-existent T-junctions are much more rare. For this reason we distinguish between the two types of junction when presenting the results. Figure 7.4 shows some of the common ways in which spurious line-endings appear.

In this experiment we consider two types of matching task. In the first of these we attempt to match lines extracted from aerial images acquired at different altitudes against a model in the form of a digital map. In this case we only anticipate feature drop-out and noise to be present in the lines segmented from the image data. We therefore operate our scheme in the scene-to-model mode. The second task is more complex and challenging; it involves the matching of the different altitude aerial images against each other. Since feature drop-out and noise will be present in both images, we operate the scene to scene method.

The experiments here involve the matching of 198 nodes in the low-level image of which 109 are T-junctions, 400 nodes in the high-level image of which 256 are T-junctions, and 158 nodes in the map of which 92 are T-junctions. Figure 7.5 shows the results of a typically matching experiment; the top set are the correct matches and the lower set are the incorrect matches. The lines representing correct matches form a conical envelope due to the

common transformation between the points in the two images, but it should be noted that our algorithm makes no use of this transformational information. The organisation comes about purely from the application of topological information to the initial label probabilities. Table 7.1 summarises the results in terms of two quantities; F_c the fraction of correct matches (correct matches/maximum available correct matches) and the noise fraction F_n (number of incorrect matches/number of non-null matches). The experiments are carried out on a variety of data sets; the full low and high altitude test images and map data and small sub-sections of these three images.

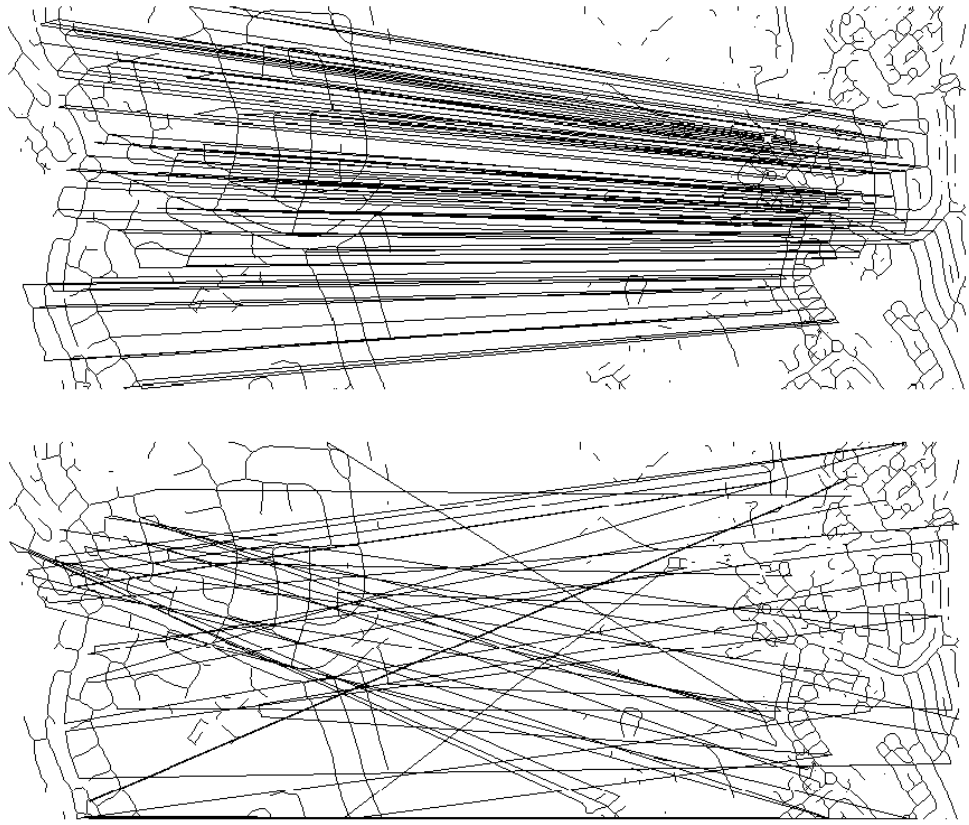


Figure 7.5: Matching results for road-network graphs using probabilistic relaxation

Table 7.1 shows the effectiveness of the method; T-junctions are matched with a correct fraction of above 80% and a noise fraction of below 16% for all data-sets except the large map to the large high image; this particular data-set represents the largest difference in scale. Because of the scale difference, there exists greater topological detail in the larger scale image than in the small scale image and a corresponding difference in the topological relationships which makes the matching a more challenging task.

Matching		R		F_c	F_n
1. small map	small low	5	T-junctions	1.0	0.0
			Line endings	0.91	0.0
2. small low	whole map	30	T-junctions	1.00	0.06
			Line-endings	0.75	0.33
3. small high	whole map	30	T-junctions	0.84	0.16
			Line-endings	0.61	0.67
4. small low	whole high	84	T-junctions	0.84	0.16
			Line-endings	0.48	0.75
5. whole low	whole map	30	T-junctions	0.81	0.16
			Line-endings	0.76	0.68
6. whole map	whole high	84	T-junctions	0.61	0.30
			Line-endings	0.64	0.64
7. whole low	whole high	84	T-junctions	0.88	0.11
			Line-endings	0.81	0.56

Table 7.1: Summary of Results from Probabilistic Relaxation

The results also illustrate the increased difficultly in matching the line-endings with this method; the results for this class are inferior to the T-junctions and the noise rejection of the algorithm is very poor.

7.8.2 Synthetic Delaunay Graphs

Finally we present the results of matching synthetically generated Delaunay graphs as described in Appendix A, using the triplet compatibility model described earlier in the chapter. Figure 7.6 summaries these results in comparison to the discrete relaxation approach. Clearly the probabilistic relaxation method is inferior in performance, however it does perform well over a wide range of corruption levels despite the lack of controlling parameters. There are two main reasons for the reduced performance as compared with our original discrete relaxation scheme; firstly although the triplet compatibilities do encourage the ordering constraint on the clique mappings, ordering can still be violated by the introduction of null matches. The discrete approach uses a dictionary of SPMs which encode *only* the legitimate mappings, so in this sense the constraints are stronger. The other factor to consider is the natural way in which the probabilistic scheme assigns corruption to a null category - the discrete scheme makes no attempt to identify noise and therefore has an easier task. The issue of identifying match noise is dealt with in Chapter 4. Both schemes converge to the same level of performance at around 60% corruption. It is also interesting to note that a fully correct match is not produced even for identical graphs - this demonstrates an inability of the scheme to recover from a poor initial labelling.

7.9 Conclusions

In this chapter we have developed a model which, in addition to the Kittler and Hancock (Kittler and Hancock, 1989) theory of probabilistic relaxation, allows us to match relational graphs based on their topology. The model so produced has proved to be entirely free from parameters and experiments have shown that this parameter free model is able to match graphs with a wide range of corruption levels without the need for any adjustments. However the scheme must be tailored to the type of graph structures expected to be present;

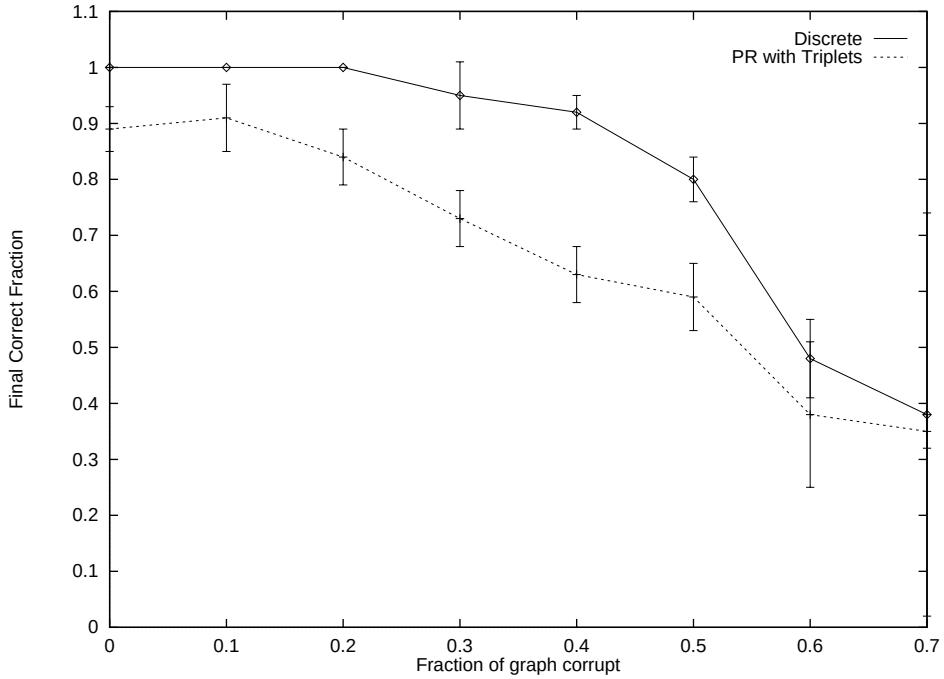


Figure 7.6: Results for synthetic graphs using probabilistic relaxation

for example a different scheme must be used when matching Voronoi triangulations and tree-like graphs. The compatibilities derived from graph topology give information concerning the power of constraints operating in the graph and the feasibility of matching different graph structures using probabilistic relaxation techniques.

The scheme is fast and converged quickly, generally in less than 10 iterations, and works well on tree-like graphs. However the triplet approach is shown to be inferior to the discrete relaxation approach on synthetic Delaunay graphs.

Chapter 8

Hierarchical Matching

8.1 Introduction

The natural world from which real vision problems are derived contains complicated structures which are difficult to model using just one type of symbolic representation or one level of abstraction. Objects more usually consist of a natural hierarchical structure in which simple scene elements such as regions of colour or texture and edges are built up into more complex entities such as surfaces, shapes and objects. In order to represent and model the hierarchical properties of real objects we need to develop methods of describing and matching hierarchical graphs.

One key element of such a method is the choice of scene primitives and groupings (Dickenson et al., 1992). This choice has implications for the modelling and matching phases of any method. The use of large and complex scene primitives means that object models can be represented by few features and the complexity of matching two such models is kept low. However the reliable recovery of complicated scene primitives is very difficult. The use of them results in a greater fraction of feature dropout and misclassification. The matching phase is then hindered by graph noise. On the other hand, simple scene primitives are easy to extract but result in a greater burden on the matching phase. For a comprehensive review of different models of object matching, see (Besl and Jain, 1985; Chin and Dyer, 1986)

Typically a hierarchical matching of scene graphs proceeds using conventional graph matching techniques such as search (Dickinson et al., 1992; Sengupta and Boyer, 1995) or

optimisation (Lau et al., 1993), with the matching results from one level being used to reduce the search space at the next level.

8.2 A Hierarchical Criterion

We begin by establishing a simple formalism to describe hierarchically structured graphs. The hierarchical structure which we describe here is based on a uniformity of object-type within a level. That is to say that all objects at one level of the hierarchy have an identical relationship to the level above, the level below and to each other. For example, corners and parallel lines could form one level of the hierarchy together, since both consist of lines at the level below, form parts of faces at the level above and may be adjacent by the sharing of scene lines. One final limitation on the hierarchy is that all objects at one level are entirely constructed from units in the level below. In other words scene detail is subsumed at the higher levels, and no new features are introduced. Figure 8.1 demonstrates an example of just such an organisation.

The hierarchy then consists of a number of levels, each containing objects which are fully described by children at the level below. Formally the levels are described by

$$G = (V^l, E^l, \forall l \in L)$$

with L being the set of levels in the hierarchy and t and b used to denote the top and bottom levels of the hierarchy respectively. V^l is the set of nodes at level l and E^l is the set of intra-level edges at level l. The nodes at level l are also characterised by a set of unary measurements denoted $\mathbf{X}^l$. The children or descendents which form the representation of an element j at a lower level are denoted by $\mathcal{D}_j$. In other words, if u^{l-1} is in $\mathcal{D}_j$ then there is a link in the hierarchy between element j at level l and element u at level $l-1$. According to our assumptions, the elements of $\mathcal{D}_j$ are drawn exclusively from V^{l-1}.

The match between scene graph G_1 and model graph G_2 is represented by a mapping function $f^l, \forall l \in L$:

$$f^l : V_1^l \rightarrow V_2^l$$

In general the upper levels of the hierarchy are more sparsely populated with entities, due to the amalgamation of scene structure into representational models. The upper levels

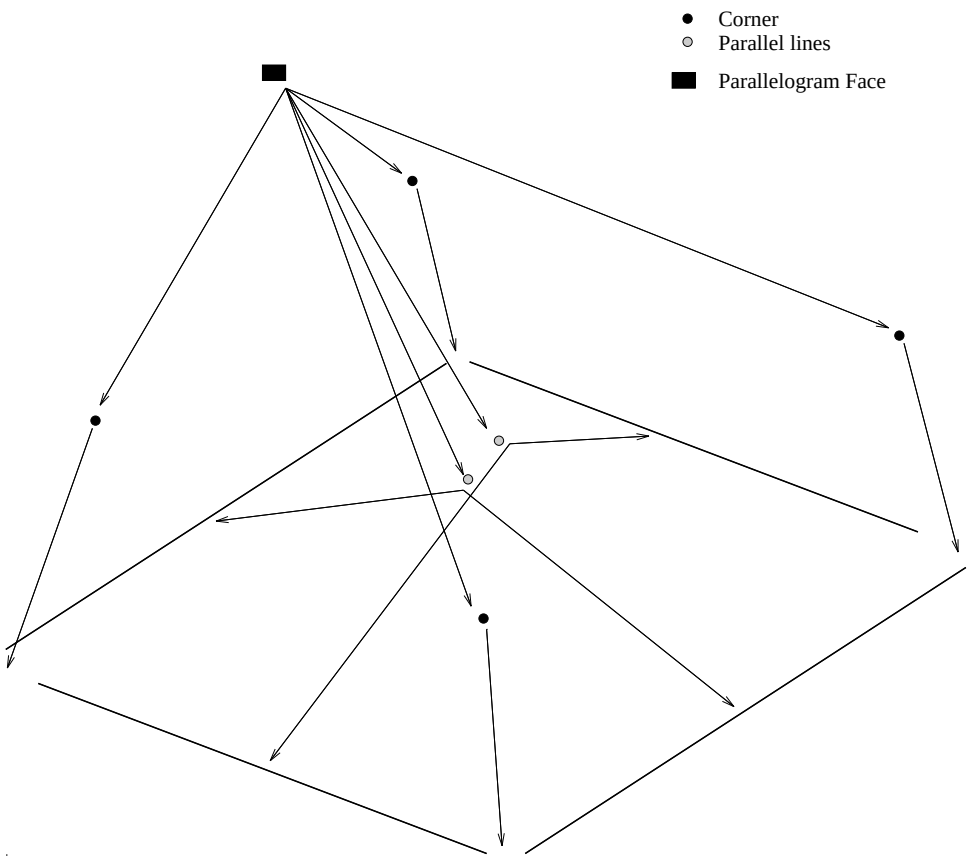


Figure 8.1: Example hierarchical graph

are more closely related to world structure while the lower levels represent image structure. It is for this reason that it is difficult to match the world-orientated upper levels using image information. Such a representational hierarchy is therefore necessary in order to propagate image information through increasingly more abstract representations.

The development of a hierarchical criterion proceeds along a similar line to the discrete criterion (Chapter 2); the quantity of interest is the MAP estimate for the mapping function f, i.e.

$$P(f^l, \forall l \in L | \mathbf{X}^l, \forall l \in L)$$

Again we are able to factorise the measurement information over the set of nodes by application of Bayes rule under the assumption of measurement independence on the nodes. The critical modelling ingredient is the joint prior of the mapping function;

$$P(f^l, \forall l \in L) \tag{8.1}$$

which represents the influence of structural information on the matching process.

The information provided to us from the scene is generally in terms of image primitives such as line segments or regions. This information is only directly relevant to the lowest level of the hierarchy in which the representation is closest to the image. Our task is therefore to propagate this information upwards through the hierarchy. To commence the formulation of a hierarchical matching scheme, we assume levels are conditionally dependent only on the levels directly above and below. This assumption allows the factorisation of the joint probability in a manner analogous to a Markov chain of probabilities. Since we wish to draw information from the bottom upwards, the chain of factorisation begins from the highest level of labelling. The expression for the joint probability of the hierarchical labelling is

$$P(f^l, \forall l \in L) = P(f^b) \prod_{l \in L, l \neq t} P(f^{l+1} | f^l) \tag{8.2}$$

We can now focus our attention on the conditional probabilities $P(f^l | f^{l-1})$. These factors express the probability of a labelling at the current level given a previously defined labelling at a lower level. We can use the concept of decomposing the graph into clique units to evaluate this probability in a similar fashion to that in Chapter 2. However in the

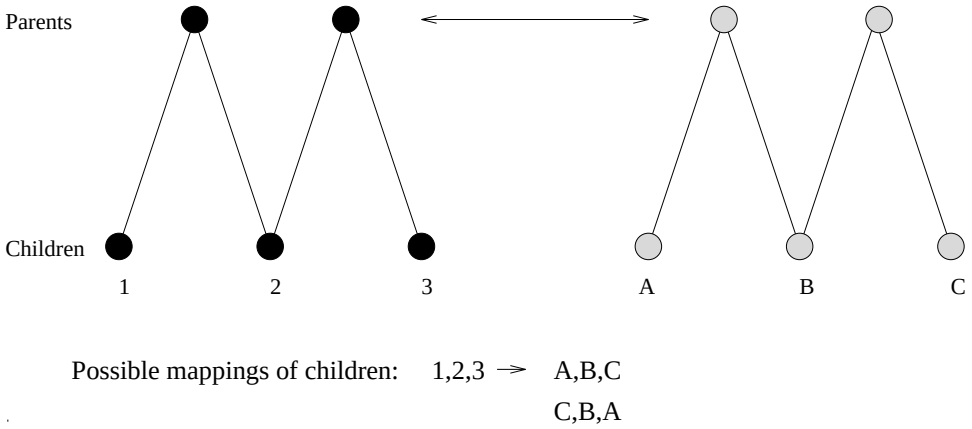


Figure 8.2: Example constrained children mappings

hierarchical case the matching of child nodes is also important in gauging the quality of match.

It is important to note that we still need to explore only those mappings which are topologically identical to the clique j and therefore the possible mappings of the child nodes are heavily constrained by the mappings of the parents (Figure 8.2)

We proceed as before with the best estimate of the conditional probability being the mean value of the clique probabilities. Therefore we write

$$P(f^l|f^{l-1}) = \frac{1}{|V^l|}\sum_{j\in V^l} P(\Gamma_j^l|f^{l-1}) \tag{8.3}$$

In order to gauge this probability, we require a dictionary of corresponding graph sub-units from G_2. These are formed in exactly the same way as the mappings were generated in Chapter 2.15; the SPMs are generated from intra-level relationships only in exactly the same fashion as with the single level criterion. We denote the set of SPMs by $\mathcal{P}$ and hence the conditional clique probability is given by

$$\begin{aligned} P(\Gamma_j^l|f^{l-1}) &= \sum_{S\in\mathcal{P}} P(\Gamma_j^l, S|f^{l-1}) \\ &= \sum_{S\in\mathcal{P}} P(\Gamma_j^l|S, f^{l-1})P(S|f^{l-1}) \end{aligned} \tag{8.4}$$

We can now see that there are two distinct elements to our model. The first element is the comparison between our mapped realisation of the clique from graph G_1, Γ_j^l, with the selected unit from graph 2 and the mapping from level $l-1$. Here we take the view that once

we have hypothesised a particular mapping from $\mathcal{P}$, the mapping f^{l-1} provides us with no further information. The matched clique Γ_j^l is conditionally independent given a mapping from the set of SPMs and we may write the first term as $P(\Gamma_j^l|S)$. The second term is the significant one in evaluating the impact of the labelling at the previous level - the possible mappings are weighted according to their probability given the mapping at the level below. The final expression for clique probability is

$$P(\Gamma_j^l|f^{l-1}) = \sum_{S\in\mathcal{P}} P(\Gamma_j^l|S)P(S|f^{l-1}) \tag{8.5}$$

All that remains now is to evaluate these two probabilities according to the node labels they contain. We represent the matched realisation of the clique in terms of node matches by $\Gamma_j^l = \{\gamma_0, \gamma_1, \ldots\}$. In this case the first term is identical to the original expression in Chapter 2 and is given by

$$P(\Gamma_j^l|S) = \prod_{\gamma_i\in\Gamma_j^l} P(\gamma_i|s_i) \tag{8.6}$$

where

$$P(\gamma_i|s_i) = \begin{cases} P_s & \text{if } \gamma_i = d \text{ or } s_i = d \\ (1-P_e)(1-P_s) & \text{if } \gamma_i = s_i \\ P_e(1-P_s) & \text{otherwise} \end{cases} \tag{8.7}$$

The second term is more subtle; it represents the conditional probability of the SPM S given a previously determined label at the level below. However the mapping contains labels only from the current level l, not labels from level $l-1$. We can reconcile this difference by noting that selection of a particular mapping at level l limits the number of consistent mappings allowed topologically at the level below. In other words if one node is mapped to another at level l, the consistent interpretation is that the children of the nodes must match to each other. By applying this constraint the labelling at $l-1$ and a set of allowed mappings of the child nodes can be used to gauge the probability of a particular SPM occuring. These legitimate mappings are referred to as Hierarchy Preserving Mappings or HPMs. We will denote the set of HPMs derived from an SPM S as $\mathcal{Q}_S$ and a member of this set as q. Using this model the conditional probability $P(S|f^{l-1})$ is given by

$$
\begin{aligned}
P(S|f^{l-1}) &= \sum_{q \in \mathcal{Q}} P(S, q|f^{l-1}) \\
&= \sum_{q \in \mathcal{Q}} P(S|q, f^{l-1}) P(q|f^{l-1}) \qquad (8.8) \\
& \qquad (8.9)
\end{aligned}
$$

We can now assume that S is conditionally independent of f^{l-1} given q, and arrive at the expression

$$P(S|f^{l-1}) = \sum_{q \in \mathcal{Q}} P(S|q) P(q|f^{l-1}) \qquad (8.10)$$

Traditionally, dictionary based hierarchical schemes have operated by using a previous labelling at another level to reduce the dictionary set by elimination of items which are inconsistent with the previous labelling. This approach can easily be incorporated into our scheme by setting $P(q|f^{l-1})$ equal to 1 for consistent items and 0 for those which are inconsistent. However we propose a different approach; by adopting the same kind of label distribution used in Equation 8.7 we can grade the SPMs according to their consistency with f^{l-1}. The model is specified by

$$P(q|f^{l-1}) = \prod_{q_i \in q} P(q_i|f^{l-1}(v_i)) \qquad (8.11)$$

where

$$P(q_i|f^{l-1}) = \begin{cases} P_s & \text{if dummy node match} \\ (1 - P_e^{l-1})(1 - P_s) & \text{if } q_i = f^{l-1}(v_i) \\ P_e^{l-1}(1 - P_s) & \text{otherwise} \end{cases} \qquad (8.12)$$

The value P_e^{l-1} must be set to reflect the prevailing level of label-errors at level $l-1$. For the conditional probability of the SPM given the HPM q, we adopt a simple uniform model under the assumption that all legitimate mappings are the same, i.e. $P(S|q) = P(S) = \frac{1}{|\mathcal{P}|}$

8.2.1 Reversibility

The expressions above have been developed under the assumption that we are ascending the hierarchy from the lowest level of scene primitives up to complete objects. Of course the choice of the lowest level as representing primitives is purely arbitrary; the consistency

criterion can also be evaluated from the top level downwards. The only change to the scheme is in the HPMs which depend on inter-level relationships. By calculating the appropriate HPMs we can traverse the hierarchy in either direction.

8.3 An Example Hierarchy

The hierarchy which we propose here is a grouping hierarchy; objects at a lower level are progressively grouped into more complex entities. There is a separation between image primitives such as line segments and regions, and more representational objects such as quadrilateral faces.

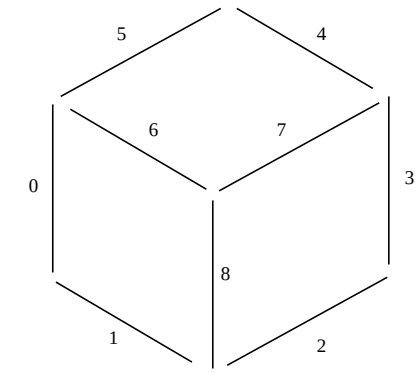
Of course the relative complexity of the various levels of grouping are of fundamental importance to the hierarchical approach both in terms of it's ability to identify meaningful structure and in terms of it's computational tractability. Simple scene objects result in a large and complex structural model because so little of the scene structure is subsumed into tokens, and correspondingly the optimisation phase of the matching has an increased burden. On the other hand, it is more difficult to reliably recover more complicated scene primitives, resulting in increasing graph noise at the more powerful representational levels (Dickenson et al., 1992).

In contrast to Dickenson et al (Dickinson et al., 1992) who use a search technique to match their hierarchical structures, we have an optimisation technique which can match large structural graphs relatively quickly. It is appropriate then to adopt relatively simple primitives and rely more heavily on the optimisation phase.

At the lowest level we adopt straight edge segments as our primitives. The natural grouping of such segments is in terms of corners and parallel lines, and these perceptual units form the second level. These groupings can be used to identify parallelograms in the scene which, under the assumption of weak perspective, represent possible rectangular object faces. These parallelograms form the top level of the hierarchy.

As an example of such a hierarchy, Figure 8.3 illustrates the hierarchical representation of one aspect of a cube in an image.

Even for a simple object such as the aspect of a cube, the hierarchical graph produced



Faces: 9,13,22,11,30,27
14,15,16,23,26,32
18,20,21,19,24,29

Corners	**9**	0	1
	10	0	5
	11	0	6
	12	1	2
	13	1	8
	14	2	8
	15	2	3
	16	3	7
	17	3	4
	18	4	5
	19	4	7
	20	5	6
	21	6	7
	22	6	8
	23	7	8
Parallelism	**24**	5	7
	25	5	2
	26	7	2
	27	1	6
	28	1	4
	29	6	4
	30	0	8
	31	0	3
	32	3	8

Figure 8.3: Hierarchical graph of a cube

is quite complex. Here we are only interested in the methodological aspects of hierarchical matching and consequently we will adopt a simpler two level hierarchy based on corner relations.

8.4 Discrete Relaxation With Hierarchical Corner Graphs

In this section we develop a hierarchical matching scheme based on line-segments and corner objects. The method of extracting these elements from the scene is explained in detail in Appendix A; here we will discuss it briefly. The lines represent either edges or lines in the image and are extracted with an edge detector (Kittler and Hancock, 1989). This edge image is then grouped into meaningful perceptual groupings using the software of Etamadi (Etemadi et al., 1991) which produces the straight-line and corner groupings of interest here.

The final element to the graph representation is a set of intra-level relationships. These relations play a parallel rôle to the edges in the non-hierarchical relaxation scheme. Again we adopt the Voronoi tessellation and associated Delaunay graph to generate the intra-level

graph edges.

8.4.1 Mappings

The structure-preserving mappings (SPMs) represent intra-level structural information and the SPMs are generated in exactly the same fashion as described in Chapter 2. The hierarchy-preserving mappings are generated in this specific corner/line representation by exploring the possible mapping of child lines given the corner match above. Since each corner consists of two line-segments, there are two HPMs for each corner participating in the SPM at the level above and hence $2|S|$ HPMs altogether.

As described earlier we can explore the hierarchy in either direction. When descending the hierarchy corner mappings become the children of lines. In this case the HPMs are determined as follows: A clique of the lines consists of a central line and the external lines which are direct Voronoi neighbours of the centre. Consider the central line and one of the external lines; if they do not mutually participate in a corner relation there is no hierarchical constraint and the hierarchical portion of probability is ignored. If on the other hand they do participate in a corner, this unit can provide a hierarchical constraint. In this case there is just one HPM; the corner must map to the corresponding corner of the mapping of the two lines (see Figure 8.4).

Armed with these SPMs and HPMs we can evaluate the hierarchical criterion in Equation 8.3. The strategy we use is to first match the lines with a non-hierarchical criterion to obtain an initial labelling. Then we match the corner level with the hierarchical criterion incorporating information from the initial line labelling. Finally the lines are re-matched with the hierarchical criterion for descending the hierarchical graph.

8.5 Experimental Results and Discussion

The discussion of the performance of the hierarchical scheme begins with a small test case to demonstrate how the hierarchical criterion can reduce ambiguity between similar graphs by introducing information from an additional level of representation. Figure 8.5 shows a graph which is symmetrical at the top level of representation and consequently ambiguous

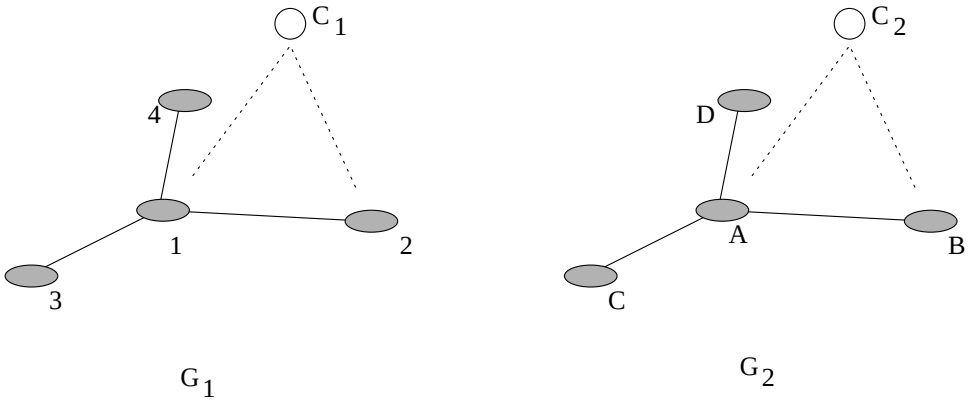


Figure 8.4: Example of an upwards HPM

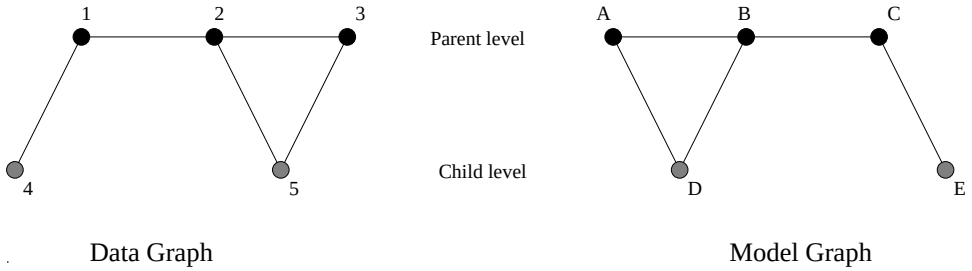


Figure 8.5: Test case: Ambiguous graphs

at that level. However at the next level there are definite differences between the children.

The graphs under consideration are ambiguous at the parent level because there is a symmetry which does not permit distinction between the mappings $(1,2,3) \rightarrow (A,B,C)$ and $(1,2,3) \rightarrow (C,B,A)$. Table 8.1 shows how the HPMs introduce information from the child level which distinguishes between the two possibilities.

However this simple case does not address some important properties of real hierarchical graphs; segmentation of real images into a hierarchical structure is difficult and prone to error. For this reason the method has also been tested on some real data; the data under study here is based on the SAR data discussed in Appendix A and consists of linear field boundaries at the lowest level and corners and the top level. Firstly a matching set of lines from the data and model has been extracted, so that all lines in the data have a matching line in the model. Corruption has then been added by deleting a certain number of lines in the data and adding the same number of lines at random positions and orientations. Figure

Figure 8.6: Example hierarchical datasets

8.6 contains an example of such a process.

Table 8.2 provides a summary of the comparison between the performance of the original discrete relaxation method on the lowest graph level of the lines, and the hierarchical approach using lines and corners. The result show that there is no improvement in the labelling with application of the hierarchical criterion, and in some cases the labelling becomes less accurate. This result can be attributed to the unreliability of the extracted perceptual relations; these relationships are unstable and corrupted to the extent that they provide dis-information to the matching at the next level.

8.6 Conclusions

In this chapter a hierarchical criterion has been developed which makes use of the concept of hierarchy preserving mapping between two hierarchical graphs. Development of this idea leads to a consistency criterion in which the probability of a particular mapping at one level is gauged by the topologically allowed mappings of the children of that mapping.

Examination of a test case has shown that this method can disambiguate graphs which are ambiguous at one level of abstraction. Results on image data with lines and corners as graph nodes expose a weakness in the hierarchical approach; the method needs reliable relational information to improve over the single level approach.

$(1,2,3) \rightarrow (A,B,C)$ HPMs	Value	$(1,2,3) \rightarrow (C,B,A)$ HPMs	Value
$5 \rightarrow D$	c	$5 \rightarrow dummy$	b
$4 \rightarrow E$	c	$dummy \rightarrow D$	b
		$4 \rightarrow dummy$	b
		$dummy \rightarrow E$	b
Total	c^2	Total	$2b^2$

Table 8.1: Mappings between the test graphs using a hierarchical criterion

Level of Corruption	Single-level relaxation F_c	Hierarchical relaxation F_c
0.15	0.5	0.45
0.2	0.67	0.56
0.27	0.47	0.47
0.34	0.42	0.42

Table 8.2: Comparison of normal and hierarchical relaxation

Chapter 9

Summary and Concluding Remarks

9.1 Summary of Contribution

In this thesis a method has been developed to match relational graphs using symbolic graph information derived from the topology of the graphs under study. The graph structures are decomposed into sub-units and it is the set of topologically valid mappings between these sub-units which provide constraint information to the matching process.

By defining a probability model in which label errors and unmatched graph nodes may appear, using a similar approach to that of Hancock and Kittler (Hancock and Kittler, 1990a), a measure of similarity between graph units is defined which takes account of some of the processes which lead to mismatches between units. This model softens the relational constraints in the sense that it tolerates a certain level of node mis-matches and missing nodes. The degree of toleration is controlled by two parameters, the probability of label error and the probability of relational corruption. As opposed to the MAP estimation scheme of Geman and Geman (Geman and Geman, 1984) these parameters have a meaning in terms of the quality of the current match and the corruption level present in the graph structures.

In contrast the attribute-based techniques of Kittler et al (Kittler et al., 1993) and Boyer and Kak (Boyer and Kak, 1988), this measure of consistency is a purely symbolic one, and in contrast to (Shapiro and Haralick, 1981) it provides a fine measure of symbolic match consistency.

This criterion is then coupled with a MAP discrete relaxation updating scheme which

makes use of a set of unary measurements on graph nodes. These unary measurements are used both to generate an initial matching state, and in the MAP update procedure.

With experimental studies on both aerial image data and synthetically generated random dot patterns, it is shown that the proposed scheme is extremely tolerant to poor initial matching conditions when the graphs are uncorrupted. For example, if the unary probabilities are sufficiently accurate to locate 10% of the correct matches initially, the relaxation scheme is able to increase this to a full 100% of matches correctly located. When the initial labelling is 5% correct, 95% of matches are located and when the initial labelling is just 2% correct, 85% of correct matches are still found. The algorithm is also extremely tolerant to occlusion of portions of the data graph, finding the fully correct match with up to 85% occlusion. Above this level however the matching is completely inoperable. The method also shows a robustness to moderate levels of corruption; with up to 20% of the nodes corrupt, 90% of the correct match is recovered. The performance degrades steadily until at 60%-70% corruption there is little advantage to the application of the relaxation scheme. Examination of the performance of a configuration-only scheme without persistent measurement information showed that such an approach is ineffective because noise from incorrect patterns interferes with the matching process.

Three methods of controlling spurious graph elements have also been studied. The first method was a constraint filtering approach applied as a post-processing step after the match has been located. This method has many similarities with maximal clique methods (Barrow and Burstall, 1976). The second method is an optimisation approach in which spurious graph nodes are labelled as such during the matching phase. Finally a new method is developed in which suspected noise elements are removed from the graph and the graph is actively reconfigured during a process which attempts to find the optimal partition between spurious nodes and valid structure. An experimental study using simulated Delaunay graphs clearly shows that the graph reconfiguration approach provides the best performance. Explicit null labelling in the optimisation phase proves to be the worst method, with constraint filtering giving the intermediate performance.

In Chapter 5 a pattern-space model was proposed for the Delaunay graph which allows the prediction of the average value of the matching criterion at a particular level of label-

error probability. Comparison with the values returned from actual runs of the algorithm demonstrated that the theoretical prediction is accurate for small values of the label error probability. A method is also proposed for using this prediction to calculate the true level of label errors at any particular stage of the matching process.

In the second part of Chapter 5, the analogy between labelling problems and statistical physics methods was exploited to develop a suitable potential for performing simulated annealing on the labelling. The entropy of the labelling was also derived.

Chapter 6 concerned itself with relating the consistency criterion developed in this thesis to some alternative methods. It was shown how the linear approximation to the criterion performs a similar operation to that of the Hopfield (Hopfield, 1984) network under certain limiting assumptions. This linear approximation was also related to a special case of the criterion of Boyer and Kak (Boyer and Kak, 1988); the condition being that Gaussian attribute deviations are used. The Boyer-Kak cost function is in this situation similar to the linear approximation of our criterion if differences in structural units are measured by the squared distance between attributes rather than Hamming distance. In a third approximate criterion the exponentials appearing in the criterion were approximated by delta functions. This approximation was shown to be equivalent to the method of Shapiro and Haralick (Shapiro and Haralick, 1981) when cliques are used as relations. A comprehensive experimental comparison of all these methods demonstrated that the exponential criterion out-performs the alternatives and exposes particular areas of weakness in the approximate approaches.

Finally in Chapter 6 a non-deterministic update method and simulated annealing were compared to the standard gradient ascent approach to optimising the matching criterion. An experimental comparison of the methods showed that there was no advantage to be gained from a more complicated optimisation method in terms of the matching performance.

Using a similar topological corruption model to that applied to the development of the discrete criterion, a scheme for the specification of compatibility coefficients has been developed within the probabilistic relaxation framework of Kittler and Hancock (Kittler and Hancock, 1989) . The compatibility coefficients in question were structurally based and entirely free from parameters, and thus were applicable under a wide range of conditions. The scheme must however be tailored to the type of neighbourhood relations present in

the graph; for example tree-like graphs used compatibilities based on node-pairs, whereas a fully triangulated planar graph such as the Delaunay graph used node-triplets. The compatibilities were also shown to provide information concerning the power of constraints operating within different graph structures. Again a set of experiments on real and synthetic datasets showed that the schemes could effectively match relational graphs, although the results were inferior to the discrete relaxation scheme.

Finally in Chapter 8 a discrete hierarchical criterion was developed using the same structural models and topology-preserving mappings which were applied to the single level case. The resulting criterion evaluated potential mappings with respect to a set of structure-preserving mappings at a single level. In turn the probabilities of these SPMs are evaluated in terms of the possible mappings of children and a previously know match at the level of the children. However the results on hierarchical image graphs were inferior to that produced by the single-level relaxation scheme.

9.2 Further Work

The control of the label-error probability is an issue which is potentially important to the performance of the discrete matching algorithm. This quantity effectively controls the level of smoothing applied to the criterion; if a dataset is under-smoothed the potential exists for the update algorithm to become trapped in local minima of the criterion. On the other hand if the function is over-smoothed, convergence to the solution becomes too slow and the update process is unnecessarily time consuming. Although in Chapter 5 a method for calculating the prevailing level of error-probability was presented, there are two problems with this approach. Firstly there is inevitable uncertainty and variation in the data to which the method is very susceptible, and therefore the resulting error-probability is unreliable. Secondly it is not clear that prevailing level of error-probability is the optimal value to use for P_e in the subsequent iteration; experimental studies suggest that this value is not optimal in terms of creating a quickly converging algorithm. It is thought that it may be possible to use a pattern space model to predict the expected depth of any local minima of the criterion and to set the label-error probability P_e accordingly.

A second unresolved issue is the poor performance of the hierarchical criterion on real hierarchical datasets. The lack of improvement with application of the hierarchical method is thought to be caused by a lack of reliable relationships in the hierarchical data. If this is the case, further work is required to provide methods for extracting reliable hierarchical relationships between elements of the scene. On the other hand, the hierarchical criterion requires careful control of the relative power of constraints from the current level and a previously labelled level. It may be this factor which is the cause of labelling problems; at this moment the issue is still unresolved.

It is also thought that the method of optimisation used during the active graph reconfiguration process (Chapter 4) has not been addressed; while we have shown that in terms of matching alone the gradient ascent method is sufficient to locate the best match, it is not clear that the method is suitable when modification of the graph structure takes place. Infact the graph modification process is rather coarse in terms of it's effect on graph structure - a node is either wholly removed or wholly present and there is no provision for softening the process in the same manner as the label-error process softens incorrect labellings. In this case a stochastic optimisation process may be more suitable, and recently considerable success has been achieved using a genetic search procedure (Cross et al., 1995). It may also be beneficial to soften graph edges and allow them to exist with a probability between 0 and 1.

Appendix A

Data Extraction and Graph Formation

A.1 Data Preparation

In order to study relational matching tasks we require data sets which are abstracted in terms of graphs. This section describes the extraction of data sets which are to be used in the experimental evaluation of the relaxation schemes. Two experimental data sets have been chosen for analysis. These two data sources are gathered using different sensors and therefore offer different challenges both in the set of features to be matched and in the different types of relational graph abstractions used to represent them. The infra-red example presented in section A.1.1 is relatively straightforward because the features are easily detected and the relationships between them clear and easily abstracted into a graph representation. In the SAR example (section A.1.2) noise and clutter are important factors and the extraction of relationships is difficult and uncertain.

A.1.1 Infra-Red Images

Figure A.1 shows a pair of infra-red(IR) line scan images. These images have a number of important features. The most salient feature of the images is the ability of the detector to pick up areas of tarmac, particularly roads. For this reason we have chosen the road network as the basic feature in these images. The images are distorted, they have a large degree of cylindrical distortion in the x-direction. The y-direction is controlled by the motion of the aircraft containing the scanner - the y-scale is dependent on the speed of that motion and

Figure A.1: Infra-red images

Figure A.2: Low and high altitude lines

distortion occurs if the aircraft changes direction.

Roads appear in the image as high-intensity line features, and are segmented using a line-finder. The line extractor, due to Kittler and Hancock (Kittler and Hancock, 1989), applies orientational line filters at four different directions in the image to enhance intensity ridges. The connectivity of the lines thus produced is enhanced by relaxation labelling using a dictionary of allowable line structures. Details of this approach are given in Ref. (Wilson and Hancock, 1993a). The output from the line detector is shown in Figure A.2.

The web of line contours is cleanly segmented from the background; there is little noise or image clutter and the contours are strongly connected. Because the line-detector dictionary explicitly encodes contour junction structure, these features are also detected with little dropout.

Figure A.3: Raw SAR Image

At the next stage a graph representation of the road network is required for relational matching to proceed. For this task I adopt an approach similar to that in (Ayache and Lustman, 1987) and (Herault et al., 1990), using the T-junctions and line-endings which delineate linear segments as feature points. These features are easily identified using the junction representation present in the line-finder. Arcs representing relationships in the graph are therefore road connections between the junctions. Measurement information is generated from the scene thus; each node is encoded with information about the line segments forming the junction which the node represents. Each line is characterised by the length of the arc between junctions and also the angle at which the line leaves the junction.

A.1.2 SAR Images

SAR(Synthetic Aperture Radar) images, in contrast to the IR data, have very little distortion of the image plane, rather they exhibit a sever noise and anisotropies associated with the directionality of the radar. The radar is sensitive changes in height at ground level - it detects elevated features such as houses, woodland and hedges. An example SAR image is shown in Figure A.3. Inspection of Figure A.3 also reveals a degree of shadowing.

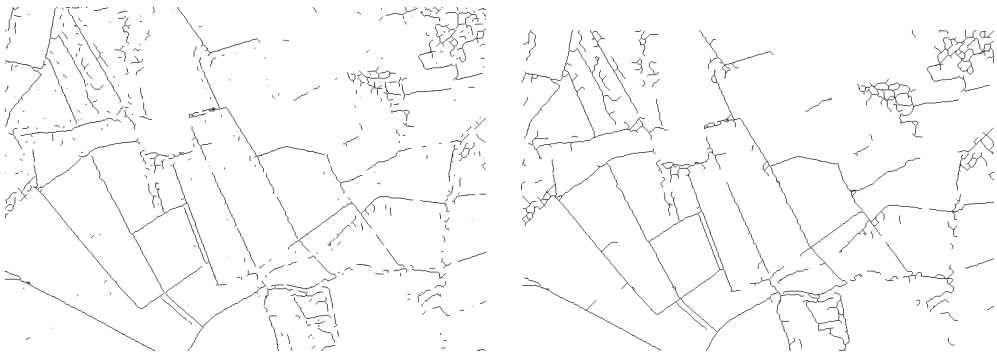


Figure A.4: Results of line detection and grouping

The most evident features in these rural scenes are the linear hedge structures along field boundaries. Again we apply the relaxational line-detector to extract these features, but with a number of refinements. The relaxational operator draws on an accurate noise model for the SAR data, due to Evans and Hancock (Evans et al., 1994). The noise distribution is specified with a Bessel function rather than the Gaussian which is used to model noise in more conventional image domains.

The extracted line contours are considerably more noisy and fragmented than in the previous IR data set. The gaps are caused by breaks in the image intensity profiles, and mirror the fragmentary nature of real hedges, which also have gaps and breaks. For our purposes we require a well connected line segment as the representation of a single continuous hedge. To this end the contour grouping of Shashua and Ullman (Shashua and Ullman, 1988) is applied to the ridge contours. The results are shown in Figure A.4.

After this processing, meaningful linear segments may be derived from the line contours. These are shown in Figure A.5.

The objective is again to form a relational graph representation of the scene, however in contrast to the IR data junctions are not well detected and lines are fragmented - the original junction/road representation will not suffice for this type of image.

Two elements are required for the formation of a relation graph. Firstly we must abstract the scene in terms of a set of 'objects' which make up the image. In Figure A.5 the image is composed of line segments - we therefore adopt these as the objects which are the nodes of the graph. The second element is a set of relationships between the objects. Ideally these relationships should as robust as possible to segmentation errors, since the matching

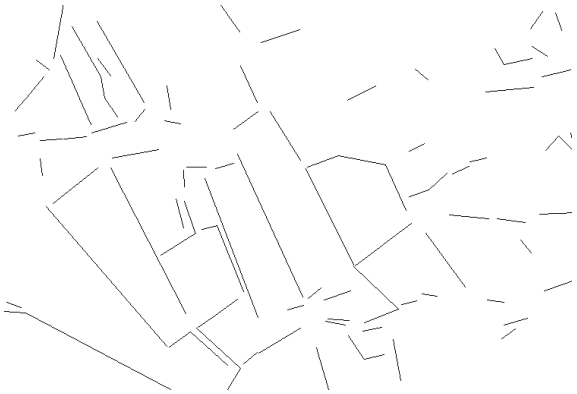


Figure A.5: Line segments

process is relation-based (Ranganath and Chipman, 1992). Geometric properties which are invariant to rotation and scale are often used (Chipman and Ranganath, 1992; Tang and Lee, 1992; Horaud and Skordas, 1989) since they have the desired robust properties over a range of image transformation. We however propose to use purely topological relations; two particular variants, Voronoi neighbours (Tuceryan and Chorzempa, 1991) and corner relations are used in this study. Topological relations have the desired invariant properties with the added advantage that their specification is purely symbolic. The choice of exactly which relation set is best is an open topic of research; I discuss the value of different relational graph structures in the thesis, within the context of relational matching tasks.

Geometric information is again encoded in the graph. In this case line lengths are considered inaccurate because of line fragmentation and are not used. Angle information is more reliable that in the IR images because of the lower level of image distortion. The measurements therefore consist of line-angles only.

A.1.3 Map data

Ordinance Survey maps of the areas detailed in the above images are also available. These are shown in Figure A.6.

These provide 'ground truth' information to which the image data can be matched. In the case of the road network the OS map provides an accurate representation of the true road network and can therefore be considered as an uncorrupted representation of the corrupted and uncertain information in the images. For the SAR data we are interested

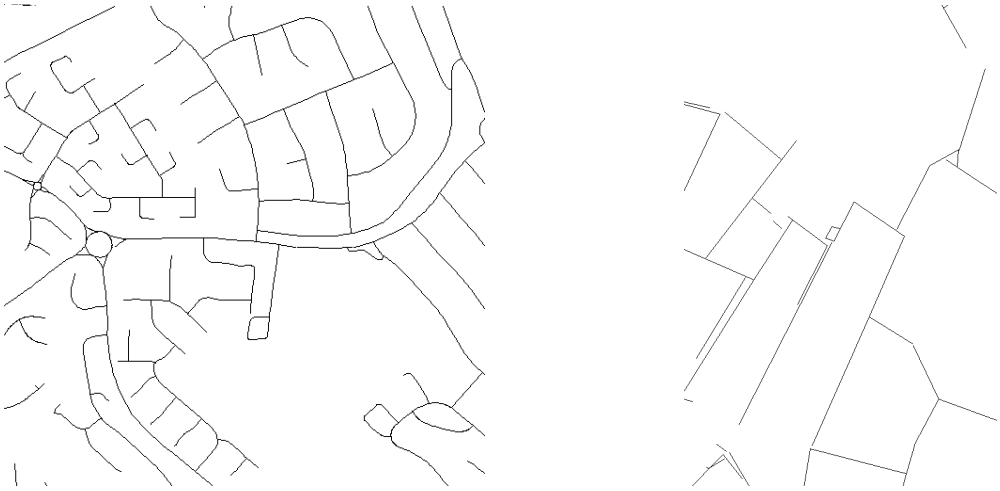


Figure A.6: Map data

in hedges - these do not appear as a separate cartographic feature in the map and so some clutter appears in the map(Figure A.6). Furthermore hedges are subject to seasonal variation and disappearance and so the map is an intrinsically inaccurate scene representation. The possibility exists of corruption in both the image and map graphs.

The maps exist in the form of vectors representing the lines present in the map. There is no need therefore for a line-detection phase; graphs are extracted in the same fashion as the corresponding image data.

A.2 Synthetic Data

While real data provides a concrete experimental test of the methods under study in this thesis, there is an insufficient quantity and variation in these data-sets to provide a rigourous examination of the performance of these matching algorithms. To achieve this end we need to generate synthetic data-sets in which we can control the levels of noise and measurement uncertainty.

The simplest and most controllable scheme we can use is the random dot pattern. Dots which represent the nodes in a graph are places at random locations in the image plan. Associated with each one of these dots is unary measurement which is used later to provide the initial matching probabilities which seed the relaxation scheme. A set of relations between these points is generated as before using the Voronoi tessellation. The output of

this process is the uncorrupted model graph.

A second pattern of dots is then generated, identical to the first aside from two alterations. The uncertainty in measurement information is incorporated by adding a Gaussian measurement error to the second measurement, the variance being under user control. To simulate corruption in the graph, a set fraction of points are removed from the graph and the same number of new points are put in their place with random locations and measurements. The Voronoi tessellation is then generated as before.

This scheme allows control over two aspects of the matching process, namely the quality of the initial match and the level of relational corruption.

A.3 Initial Match Probabilities

The essence of the relaxation approach is to refine a set of initial labellings or matches by incorporating information from neighbouring matches. In this fashion consistency can be imposed on the labeling. An initial labelling is required in order to seed the relaxation scheme; in the case of graph-matching these take the form of a set of initial, non-contextual match probabilities between nodes in the two graphs.

The modelling of the initial probabilities draws on transformational differences between the scenes under study, and is based on the geometric properties of the lines forming those scenes. Both segmental inaccuracies and geometric distortion are present and so these must be captured by the probability distributions proposed to explain the measurements.

The simplest case is that of the SAR images. Lines in these images are encoded only with unary angle information. Furthermore line angle is unaffected by segmentation errors such as line fragmentation. We need only account for uncertainties present in the process of angle detection. The angle affinity between two lines is therefore given by the following Gaussian distribution

$$\rho_\theta(u,v) = \exp\left[-\frac{(\theta_{u,v}-\Theta)^2}{2\sigma_\theta^2}\right] \tag{A.1}$$

where $\theta_{u,v}$ is the angle between lines u and v, Θ is the relative rotation of the two images, and σ_θ^2 is the variance anticipated in the angle. Under the assumption of a uniform prior

distribution of possible line matches, the match probability becomes

$$P^{(0)}(u,v) = \frac{\rho_\theta(u,v)}{\sum_{u\in G_1}\sum_{v\in G_2}\rho_\theta(u,v)} \tag{A.2}$$

In the case of the junction matching problem presented by the IR images we need to combine evidence from all the lines that form a junction. We have both angle and length information for these lines and so have the added complexity of incorporating line-lengths which are subject to segmentation error. In the graph description the line-sets making up junction u are given by $\mathcal{L}_u = \{e_1 | u \in e_1\}$.

Given two sets of arcs $\mathcal{L}_u$ and $\mathcal{L}_v$ we must consider the support these arcs offer to a junction match. Since we do not know which arc matches to which we must compile support over all possible match combinations; the probability of a junction match is then given by

$$P^{(0)}(u \to v) = \frac{\prod_{e_1\in L_u}\sum_{e_2\in L_v} q(e_1 \to e_2)}{\sum_{w\in V_2}\prod_{e_1\in L_u}\sum_{e_2\in L_w} q(e_1 \to e_2)} \tag{A.3}$$

where $q(e_1 \to e_2)$ is the support for a match between arcs e_1 and e_2. As before the support based on angle is given by ρ_θ. For lengths the support is more complex; we begin by postulating another exponential distribution in the absence of segmentation error

$$\rho_s(e_1,e_2) = \exp\left[-\frac{(S_{e_1,e_2}-S)^2}{2\sigma_s^2}\right] \tag{A.4}$$

where S is the relative scaling of the images and $S_{e_1,e_2} = l(e_1)/l(e_2)$ is the relative scale between the candidate lines. The possibility of segmentation error is modelled by assuming a uniform probability of line corruption p. The support is then summed over the possibilities of correct segmentation and segmentation error

$$q(e_1 \to e_2) = \Big\{(1-p)\rho_s(e_1,e_2) + p(1-\rho_s(e_1,e_2))\Big\}\rho_\theta(e_1,e_2) \tag{A.5}$$

Finally the synthetic data is seeded simply using the probabilities generated from the affinity

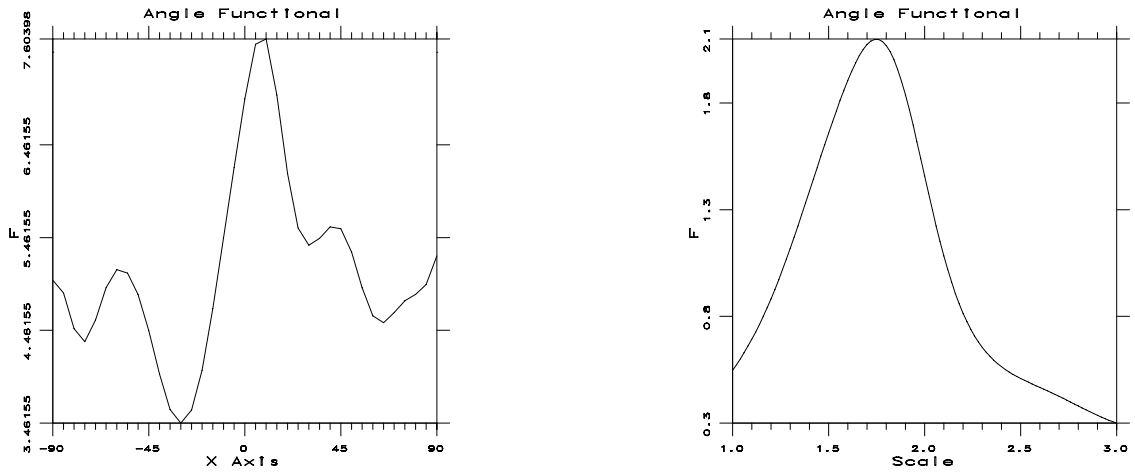


Figure A.7: Initial Probability criterion

$$\rho(u,v) = \exp\left[-\frac{(\mathbf{x}_u - \mathbf{x}_v)^2}{2\sigma^2}\right] \tag{A.6}$$

Null labels are given an initial probability equal to the null label probability i.e. $P(f(u) = \phi) = P_n$

With the synthetic data, the variance of the measurements is already known to be that which was originally used to generate the data. For the other datasets determining the transformation and variance parameters is more of a problem. In the experimental studies here, these parameters have been determined using a sample of correct matches from a human observer. However the parameters can be determined approximately in an automated way; we first define an initial match criterion which is the sum of the initial match affinities

$$\mathcal{F} = \sum_{u \in V_1} \sum_{v \in V_2} \rho(u,v) \tag{A.7}$$

Example plots of this functional are shown in Figure A.7 when and S are varied for the initial probabilities of the low road data to the high data. The correct orientation and scale correspond to global maxima of the criterion. However in the plot of orientation there are a number of local maxima; these correspond to rotational symmetries of many of the road junctions. The variance in the angle can be roughly estimated from the width of the principle maxima.

A.4 Experimental Protocol

During experimental runs, unless otherwise stated, parameters of the relaxation schemes are set as follows:

- 10 iterations are used for the probabilistic relaxation scheme, 14 for the discrete approach. It should be noted that these values are rather higher that is usually necessary for convergence, but are set thus to guarantee that the scheme has fully converged.
- The null label probability, where used is set as follows; 0.12 for the road network data, 0.5 for the SAR data and equal to the level of corruption for the synthetic data.
- The label error probability in the discrete relaxation algorithm exponentially decays from an initial value of 0.3 to 0.0003 on the final iteration

With the synthetic data, four experiments are run at each level of corruption, the matching performance being the mean of the runs. The standard deviation provides the error-bars on plots of synthetic results.

Matching performance is characterised by the following measure

$$F_c = \frac{N_c}{N_p}$$

where N_c is the number of correct matches which are found and N_p is the possible number of correct matches available. Clearly the algorithm cannot find more matches than the number of uncorrupted nodes in the graph; using this definition the matching performance is represented by an number from 0 to 1 for all matching problems, making them directly comparable. The ability of the algorithm to reject matching noise is characterised by

$$F_n = \frac{N_w}{N}$$

where N_w is the number of incorrect matches which are found and N is the number of matches produced. This quantity measures how much noise is present in the match. Again the measure ranges from 0.0 to 1.0, however in this case 0.0 is the best performance (no incorrect matches present).

The numbers of correct and incorrect matches are calculated by comparison with a 'ground truth' match. In the case of synthetic data, the correct matches are already known. The set of correct matches for the real data must be generated by hand.

Bibliography

Ahuja, N., An, B., and Schachter, B. (1985). Image representation using the Voronoi tessellation. *CVGIP*, 29:286–295.

Ayache, N. and Lustman, F. (1987). Efficient registration of stereo images by matching graph descriptions of edge segments. *Int. J. of Computer Vision*, 1:107–131.

Ballard, D. H. and Brown, C. H. (1982). *Computer Vision*. Prentice-Hall.

Barrow, H. G. and Burstall, R. M. (1976). Subgraph isomorphism, matching relational structures and maximal cliques. *Information Processing Letters*, 4:83–84.

Barrow, H. G. and Popplestone, R. J. (1971). Relational descriptions in picture processing. *Machine Intelligence 6: Edinburgh*.

Besl, P. J. and Jain, R. C. (1985). Three-dimensional object recognition. *ACM Computing Surveys*, 17(1):75–14.

Bhanu, B. and Faugeras, O. D. (1984). Shape matching of two-dimensional objects. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 6(2):137–155.

Boyer, K. and Kak, A. (March 1988). Structural stereopsis for 3-D vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10 A02(2):144–166.

Chin, R. T. and Dyer, C. R. (1986). Model-based recognition in robot vision. *ACM Computing Surveys*, 18(1):67–108.

Chipman, L. J. and Ranganath, H. S. (1992). A fuzzy relaxation algorithm for matching imperfectly segmented images to models. *Proc. IEEE South East Conference*, pages 128–136.

Clowes, M. (1971). On seeing things. *Artificial Intelligence*, 2:79–116.

Cohen, F. and Cooper, D. (1987). simple parallel hierarchical and relaxation algorithms for sementing non-causal Markovian random fields. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 9:195–219.

Cross, A. D. J., Wilson, R. C., and Hancock, E. R. (1995). Discrete relaxation on a boltzmann machine. *Proceedings of the 1995 International Conference on Artificial Neural Networks*, pages 491–496.

Davis, L. S. (1979). Shape matching using relaxation operators. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 1:60–72.

Dickenson, S. J., Pentland, A. P., and Rosenfeld, A. (1992). From volumes to views: An approach to 3-D object recognition. *CVGIP: Image Understanding*, 55(2):130–154.

Dickinson, S. J., Pentland, A. P., and Rosenfeld, A. (1992). 3-D shape recovery using distributed aspect matching. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 14(2):174–197.

Etemadi, A., Schmidt, J.-P., Matas, G., Illingworth, J., and Kittler, J. (1991). Low-level grouping of straight line segments. *Proc. of British Machine Vision Conference*, pages 118–126.

Evans, A., Sharp, N. G., and Hancock, E. R. (1994). Noise models for linear feature detection in SAR images. *Int. Conf. on Image Processing*, 1:466–470.

Faugeras, O. D. (1981). Decomposition and decentralization techniques in relaxation labelling. *Computer Vision and Image Processing*, 16:341–355.

Faugeras, O. D. and Berthod, M. (1981). Improving consistancy and reducing ambiguity in stochastic labelling: An optimisation approach. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 3:412–424.

Finch, A. M., Wilson, R. C., and Hancock, E. R. (1995). Matching delaunay triangulations by probabilistic relaxation. *Lecture Notes in Computer Science, Springer-Verlag*, 970:350–358.

Flynn, P. and Jain, A. (Oct. 1991). Bonsai: 3D object recognition using constrained search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1066–75.

Fogel, D. B. (1994). An introduction to simulated evolutionary optimisation. *IEEE Transactions on Neural Networks*, 5:3–14.

Gardner, E. (1986). Structure of metastable states in the Hopfield model. *Journal of Physics A*, 19(16):L1047–L1052.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and Bayesian restoration. *IEEE Transations Pattern Analysis and Machine Intelligence*, 6:721–741.

Gidas, B. (1989). A renormalization group approach to image processing problems. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 11:164–180.

Hancock, E. R., Evans, A. N., and Wilson, R. C. (1994). Segmenting and matching sar images by relaxation labelling. *Workshop on SAR Image Segmentation*, pages 30–37.

Hancock, E. R., Haindl, M., and Kittler, J. (1992). Multiresolution edge labelling using hierarchical relaxation. 11^{th} *International Conference on Pattern Recognition*, 2:140–144.

Hancock, E. R. and Kittler, J. (1990a). Discrete relaxation. *Pattern Recognition*, 23:711–733.

Hancock, E. R. and Kittler, J. (1990b). Edge labelling using dictionary-based relaxation. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 12:165–181.

Hancock, E. R. and Kittler, J. (1993). A Bayesian interpretation of the Hopfield network. *IEEE International Conference on Neural Networks*, 1:341–346.

Hancock, E. R. and Wilson, R. C. (1995). Rectifying structural matching errors. *Recent Progress in Computer Vision, Springer-Verlag*, 1035.

Harary, F., editor (1969). *Proof Techniques in Graph Theory*. Addison-Wesley, Reading, MA.

Henderson, T. C. (1990). *Discrete Relaxation Techniques*. Oxford University Press.

Herault, L., Horaud, R., Veillon, F., and Neiz, J.-J. (1990). Symbolic image matching by simulated annealing. *Proc. British Machine Vision Conference*, pages 319–400.

Hopfield, J. (1984). Neural networks and physical systems with emergent collective computational capabilities. *Proc. Natl. Acad. Sci.*, 79:2554–2558.

Horaud, R. and Skordas, T. (1989). Stereo correspondance through feature grouping and maximal cliques. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 11:1168–1179.

Horaud, R., Veillon, F., and Skordas, T. (1990). Finding geometric and relational structures in an image. *Proc. of 1st Euro. Conf. on Computer Vision*, pages 374–384.

Huffman, D. A. (1971). Impossible objects as nonsense sentences. *Machine Intelligence*, 6:295–323.

Hummel, R. A. and Zucker, S. W. (1983). On the foundations of relaxation labelling processes. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 5:267–287.

Izumi, M., Asano, T., Fukunaga, K., and Murata, H. (1992). Matching of edge-line images using relaxation. *IEICE Trans. Inf. and Syst.*, E75-D(6):902–908.

Jones, G. and Wong, K. C. Heirarchical stereopsis of perceptual groups in an optimisation framework. *Progress in Image Analysis and Processing II*.

Kittler, J., Christmas, W. J., and Petrou, M. (1993). Probabilistic relaxation for matching problems in machine vision. *Proc. 4th International Conf. on Computer Vision*, pages 666–674.

Kittler, J. and Hancock, E. R. (1989). Combining evidence in probabilistic relaxation. *Int. J. of Pattern Recognition and Artificial Intelligence*, 3:29–51.

Kittler, J. and Taylor, C. J. (1994). Statistical methods in vision. *BMVC Tutorial*.

Kullback, S. (1987). The Kullback-Leibler distance. *American Statistician*, 41(4):340.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86.

Lau, W. H., Hancock, E. R., and Wilson, R. C. (1993). Hierarchical relaxation. *Proc. of the 5th University of New Brunswick AI Symposium*, pages 111–121.

Li, S. Z. (1992). Matching:invariant to translations, rotations and scale changes. *Pattern Recogntion*, 25:583–594.

Lloyd, S. A. (1983). An optimisation approach to relaxation labelling algorithms. *Image and Vision Computing*, 1(2):85–91.

Lu, Y. and Jain, R. (1992). Reasoning about edges in scale-space. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 14:450–468.

Luo, R. C., Lin, and Scherp (1988). Dynamic multi-sensor data fusion system for intelligent robots. *IEEE J. of Robotics and Auto.*, 4.

Marr, D. (1984). *Vision*. W. H. Freeman and Co., San Francisco.

Meer, P., Sher, C., and Rosenfeld, A. (1990). The chain-pyramid: Hierarchical contour processing. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 12:363–376.

Messmer, B. T. and Bunke, H. (1994). Efficient error-tolerant sub-graph isomorphism detection. *Shape, Structure and Pattern Recognition, Ed. D. Dori and A. Bruckstein, World Scientific*.

Milun, D. and Sher, D. (1993). Improving sampled probability distributions for markov random fields. *Pattern Recognition Letters*, 14:781–788.

Mjolsness, E., Gindi, G., and Anandan, P. (1989). Optimisation in model matching and perceptual organisation. *Neural Computation*, 1(2):218–229.

Price, K. E. (1985). Relaxation matching techniques - a comparison. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 7(5):617–623.

Ranganath, H. S. and Chipman, L. J. (1992). A fuzzy relaxation approach for inexact scene matching. *Image and Vision Computing*, 10(9):631–640.

Rose, K., Gurewitz, E., and Fox, G. C. (1993). Constrained clustering as an optimisation method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:785–794.

Rosenfeld, A., Hummel, R. A., and Zucker, S. W. (1976). Scene labelling by relaxation operations. *IEEE Transactions on Systems, Man and Cybernetics*, 6:420–433.

Sanfeliu, A. and Fu, K. S. (1983). A distance measure between attributed relational graphs for pattern-recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 13(3):353–362.

Sarkar, S. and Boyer, K. L. (1993). Preceptual organization in computer vision. *IEEE Trans. Systems, Man and Cybernetics*, 23:382–399.

Sengupta, K. and Boyer, K. L. (1995). Organizing large structural modelbases. *IEEE Pattern Analysis and Machine Intelligence*, 17(4):321–332.

Shapiro, L. G. and Haralick, R. M. (1981). Structural descriptions and inexact matching. *IEEE Transations on Pattern Analysis and Machine Intelligence*, 3(5):504–519.

Shapiro, L. G. and Haralick, R. M. (1985). A metric for comparing relational descriptions. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 7(1):90–94.

Shashua, A. and Ullman, S. (1988). Structural saliency: the detection of globally salient structures using a locally connected network. *Proc. of 2nd International Conf. on Computer Vision*, pages 321–327.

Tang, Y. C. and Lee, C. S. G. (1992). A geometric feature relation graph formulation for consistant sensor fusion. *IEEE Transactions on Systems, Man and Cybernetics.*, 22:115–129.

Ton, J. and Jain, A. K. (1989). Registering landsat images by point matching. *IEEE Transations on Geoscience and Remote Sensing*, 27:642–651.

Tsai, W. H. and Fu, K. S. (1983). Subgraph error-correcting isomorphisms for syntactic pattern-recognition. *IEEE Transations on Systems, Man and Cybernetics*, 13(1):48–62.

Tuceryan, M. and Chorzempa, T. (1991). Relative sensitivity of a family of closest-point graphs in computer vision applications. *Pattern Recognition*, 24(5):361–73.

Ullman, J. R. (1976). An algorithm for sub-graph isomorphism. *Journal ACM*, (23):504–519.

Waltz, D. L. (1975). *Understanding line drawings of scenes with shadows*. McGraw-Hill, New York.

Wang, C. and Abe, K. (1995). Region correspondence by inexact attributed planar graph matching. *Proceedings of the International Conference on Computer Vision 1995*, pages 440–447.

Wilson, R. and Hancock, E. R. (1994a). Matching features in aerial images by relaxation labelling. *Progress in Image Analysis and Processing III*, pages 209–217.

Wilson, R. C., Evans, A. N., and Hancock, E. R. (1994). Relational matching by discrete relaxation. *Proceedings of the Fifth British Machine Vision Conference*, pages 43–54.

Wilson, R. C., Evans, A. N., and Hancock, E. R. (1995). Relational matching by discrete relaxation. *Image and Vision Computing*, 13:411–422.

Wilson, R. C. and Hancock, E. R. (1993a). Relaxation matching of road networks in aerial images using topological constraints. *Proc. of SPIE Sensor Fusion VI*, pages 444–455.

Wilson, R. C. and Hancock, E. R. (1993b). A topological constraint corruption process for graph matching. *Proc. of the Czech Pattern Recognition Workshop*, pages 18–25.

Wilson, R. C. and Hancock, E. R. (1994b). Compatibility modelling for graph matching. *Aspects of Visual Form Processing*, pages 574–584.

Wilson, R. C. and Hancock, E. R. (1994c). Graph matching by configurational relaxation. *Proceeding of the 12th Internation Conference on Pattern Recognition*, pages 563–566.

Wilson, R. C. and Hancock, E. R. (1994d). Graph matching by discrete relaxation. *Pattern Recognition in Practice IV*, pages 165–176.

Wilson, R. C. and Hancock, E. R. (1995a). Inexact graph matching criteria. *Shape, Structure and Pattern Recognition*, pages 251–260.

Wilson, R. C. and Hancock, E. R. (1995b). An integrated approach to grouping and matching. *Lecture Notes in Computer Science, Springer-Verlag*, 974:62–67.

Wilson, R. C. and Hancock, E. R. (1995c). Relational matching with active graphs. *Lecture Notes in Computer Science, Springer-Verlag*, 970:334–341.

Wilson, R. C. and Hancock, E. R. (1995d). Relational matching with dynamic graph structures. *Proceedings of the Fifth International Conference on Computer Vision*, pages 450–456.

Wong, A. K. C. and You, M. (1985). Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:509–609.

Yuille, A. (1990). Generalized deformable models, statistical physics and matching problems. *Neural Compution*, 2:1–24.

Ziqing Li, S. (1990). Towards 3D vision from range images: An optimisation framework and parallel distributed networks. *Ph.D. Thesis, University of Surrey*.

Zucker, S. W. and Mohammed, J. L. (1978). Analysis of probabilistic labeling processes. *Proc. IEEE Pattern Recognition In Progress Conf. Chicago, IL, USA*, pages 307–312.