Relaxation labelling algorithms — a review

J Kittler and J Illingworth

An important research topic in image processing and image interpretation methodology is the development of methods to incorporate contextual information into the interpretation of objects. Over the last decade, relaxation labelling has been a useful and much studied approach to this problem. It is an attractive technique because it is highly parallel, involving the propagation of local information via iterative processing. The paper surveys the literature pertaining to relaxation labelling and highlights the important theoretical advances and the interesting applications for which it has proven useful.

Keywords: image processing, image interpretation, relaxation labelling

Many sensory data understanding problems involve the task of detecting and identifying objects the data represents. For instance, in continuous speech recognition the sensory data is an acoustic signal, and objects to be detected and identified may be words or phonemes. The richest form of sensory data is the image. Objects at a high level of description will be the elements of an imaged scene while at lower levels objects will be edges, image segments, simple geometric shapes etc.

The identification of each object is usually based on a set of measurements which can be extracted from the sensory data and which characterize and specifically relate only to one object. For instance, in the case of edges the measurements may be local directional derivatives of the image intensity function. In shape recognition, the set of measurements may comprise the Fourier descriptors of the object boundary.

Such object-specific information is seldom sufficient to allow unambiguous interpretation of the object. Nevertheless, as collections of objects in the real world are invariably ordered in some sense, the constraints imposed by the mutual relationships between individual objects in ordered collections may be used to reduce or even eliminate the ambiguity. The idea can be best exemplified by the properties of collections of objects in natural speech where words in a sentence are subject to grammatical rules of the language. These rules provide a powerful source of information which may be adequate to eradicate any errors that might otherwise be made if word recognition were based simply on the properties of the acoustic signal representing each word.

Depending on the nature of the rules which describe the relationships between objects, the presence of a particular object in a collection may impose constraints on certain classes of objects in its neighbourhood and by the same token may actively provide support for others. In general, therefore, irrespective of whether it has an inhibiting or supportive role, each object can be considered as a source of contextual information about other objects and may therefore aid their interpretation.

The idea of using contextual information is far from new. Ullmann\(^1\) exploited constraints imposed by triplets of pattern primitives to reduce substantially the errors that occur with a pattern recognition system after a learning sequence of fixed length. Clowes\(^2\) and Huffman\(^3\) used constraints between straight line segments to eliminate nonsensical interpretations of an ideal line drawing representing a set of polyhedra. Many other methods of using contextual information have been suggested in the literature (eg by Zucker\(^4\), Kittler\(^5\) and Foglein\(^6\) and Haralick\(^7\)). They broadly fall into the following categories: group classification methods, contextual decision rules, rewriting rules (grammatical or structural methods) and relaxation methods. This paper will concentrate on the last category, namely on relaxation labelling methods in general and on probabilistic relaxation labelling in particular. The discussion will be restricted to imagery data but generally any method suggested for two-dimensional image data is equally applicable to one-dimensional sensory signals.

The pioneering work in relaxation labelling is normally credited to Waltz\(^8\) who considered the problem of line drawing interpretation studied earlier by Clowes\(^2\) and Huffman\(^3\). His formulation of the consistent labelling problem allowed only unambiguous interpretation of line segments. This was achieved by
sequentially filtering out inconsistent label pairs on connected segments. This approach was popularized by Rosenfeld et al. who showed that Waltz’s filtering can be carried out in parallel and could therefore be implemented as a network of processors, each associated with one object in the image. An extensive theoretical underpinning of the consistent labelling problem has been given by Montanari, Mackworth, Haralick, Henderson, Ullmann et al., Shapiro, Nudel, Nishiura and Ikeda, and Kasif and Rosenfeld.

However, the problem considered by Waltz is somewhat unrealistic since no information as to the identity of each line segment is assumed to be available. In practice, when analysing real imagery, it is reasonable to assume that some useful information could be extracted from the raw image data. On the other hand, the edge representation of a scene is unlikely to look like an ideal line drawing. Rosenfeld et al. argued that, perhaps more appropriate than discrete relaxation which enforces unambiguous labelling, it is better to formulate the problem in a continuous domain. Fuzzy set and probabilistic frameworks were considered in this respect but the latter seems to have attracted most attention to date.

This paper reviews the literature on probabilistic relaxation now comprising more than 100 contributions. The last major reviews of this subject area, which were carried out by Davis and Rosenfeld, Faugeras, Berthod and Ballard and Brown, date from 1981 and by now the number of papers written on the topic has more than doubled.

In the next section, the classical probabilistic relaxation labelling algorithm of Rosenfeld et al. will be overviewed. The three sections after that summarize theoretical and methodological developments of probabilistic relaxation. The material is discussed under three major headings; probability updating schemes, compatibility and support functions, and general issues which do not fall neatly under any of the previous two headings. A discussion of applications of relaxation is followed by some brief conclusions.

It should be noted that relaxation processes are widely used for solving problems of the image filtering type, where the task consists in recovering a function defined on the image domain from the observed image intensity values. Although such a task could be formulated as a labelling problem, each grey level could be associated with a different label, its solution using probabilistic labelling methodology would be unnatural and inefficient. An excellent survey of many of these approaches has been given by Glazer.

CLASSICAL PROBABILISTIC RELAXATION

Before reviewing the literature on probabilistic relaxation, it is best to give a brief overview of the classical probabilistic relaxation labelling method introduced in the seminal paper of Rosenfeld et al. This widely cited and used scheme will provide a reference point to which subsequent developments of the methodology can be conveniently related.

Consider a network of objects $a_i, i = 1, 2, \ldots, N$. Unit $a_i$ can belong to one of $m$ possible classes from the set $\Lambda$. For the sake of simplicity we shall assume that the label sets $\Lambda_i$ are identical for all the nodes of the network, i.e.

$$\Lambda_i = \Lambda = \{\lambda_k | k = 1, \ldots, m\} \quad \forall i$$

Each unit $a_i$ will usually be characterized by a set of measurements $X_i$. In addition, all the objects in the network will provide contextual information which may be relevant to the interpretation of objects $a_i$. The task addressed by relaxation labelling is to assign labels $\theta_i$ to objects $a_i, i = 1, \ldots, N$ taking into account the total evidence provided by measurements $X_i$ and the contextual information conveyed by the network.

Ideally we would like to choose for each node $i$ just one label from the set $\Lambda$, thus giving an unambiguous labelling to the whole network. However, in practice the available information will be insufficient to make such unambiguous interpretation possible. Instead we shall have to settle for a probabilistic result obtained by allowing all the possible labels $\lambda_k$ for every node $a_i$ and by assigning probability value $P(\theta_i = \lambda_k | X_i)$ contextual information to each of these labels. In the framework of this generalized formulation the unambiguous labelling of object $a_i$ becomes a special case of a probabilistic labelling corresponding to the situation when the probability for one of these labels equals unity and that of the other labels assumes the value zero.

To describe the relaxation labelling scheme, it is first necessary to introduce essential notation. The contextual information conveyed by label $\lambda_k$ at object $a_i$ about label $\lambda_j$ at object $a_j$ will be represented by a compatibility coefficient $r(\theta_i = \lambda_k, \theta_j = \lambda_j)$. Let the current (s-th) estimate of the probability that node $a_i$ has label $\lambda_k$ be $P^s(\theta_i = \lambda_k)$. Then the support for label $\lambda_k$ at $a_i$, given by node $j$ can be defined as

$$q_j^s(\theta_i = \lambda_k) = \sum_{k=1}^{m} r(\theta_i = \lambda_k, \theta_j = \lambda_j) P^s(\theta_j = \lambda_j) \quad (2)$$

and the total support for this label lent by all the objects in the network is defined as

$$Q_j^s(\theta_i = \lambda_k) = \sum_{j=1}^{N} C_j q_j^s(\theta_i = \lambda_k) = \sum_{j=1}^{N} \sum_{k=1}^{m} r(\theta_i = \lambda_k, \theta_j = \lambda_j) P^s(\theta_j = \lambda_j) \quad (3)$$

where $C_j$ are weights satisfying

$$\sum_{j=1}^{N} C_j = 1 \quad (4)$$

The coefficients $C_j$ can be used to express the relative importance assigned to individual nodes in the network.

With the basic notation introduced it is now possible to write down the expression suggested by Rosenfeld et al. for computing updated probabilities of object labels. Given the current value of probability $P^s(\theta_i = \lambda_k)$, the (s+1)th estimate $P^{s+1}(\theta_i = \lambda_k)$ is defined as

$$P^{s+1}(\theta_i = \lambda_k) = \frac{P^s(\theta_i = \lambda_k)}{\sum_{k=1}^{m} P^s(\theta_i = \lambda_k)} Q_j^s(\theta_i = \lambda_k) \quad (5)$$

where the normalization term in the denominator...
guarantees that the updated quantities satisfy the axiomatic properties of probabilities.

The probability updating formula in equation (5) is largely intuitive and its basis is that the current label probabilities at each node should be adjusted according to the relative support afforded to the different labels by the complete network. It is often referred to as the nonlinear probabilistic relaxation scheme. Its linear counterpart, also suggested by Rosenfeld et al.69, aimed to find label probabilities satisfying

\[ P(\theta = \lambda_i) = Q(\theta = \lambda_i) \quad \forall i,j \]

(6)

However, the linear relaxation has been shown to yield a unique solution which is independent of any relevant information that may be contained in measurements \( X_i \) unless the formulation is modified as suggested by Blake99 and Elfving and Eklundh65. In general, this information is incorporated in the relaxation process by means of initial probabilities \( P^0(\theta = \lambda_i) \quad \forall i,j \) which are determined as

\[ P^0(\theta = \lambda_i) = P(\theta = \lambda_i | X_i) \]

(7)

where the right-hand side in equation (7) denotes the \textit{a posteriori} probability of label \( \theta \) assuming the value \( \lambda_i \) given the measurement set \( X_i \). It is, in fact, the ‘no-context’ probability of the \( i \)th node having label \( \lambda_i \). Note that this information in \( X_i \) is used directly only once at the algorithm initialization step.

The algorithm in equation (5) generated considerable interest which is partly reflected in the number of its applications reported in the literature. These will be reviewed in some detail below. First the following section summarizes the basic properties of the algorithm which have been observed by various researchers. This is followed by a survey of the theoretical and methodological developments which have taken place to date.

PROBABILITY UPDATING SCHEMES

While the probabilistic relaxation scheme in equation (5) has been demonstrated in several examples to have a tendency to increase the consistency of labelling of objects in a given relational network, its formal convergence properties were not established by Rosenfeld et al.69. Moreover, it has been observed that the convergence of the algorithm can be slow.62 The first attempt to fill this gap in the understanding of these properties was reported by Zucker et al.69. However, the condition of convergence put forth in their paper turned out to be necessary but not sufficient. In any case the paper contained no suggestion as to how the convergence condition could be translated into conditions on the actual compatibility and support functions to ensure convergence of the updating scheme. The problem of slow convergence was dealt with in a heuristic fashion by effectively taking the right-hand side of equation (5) to the power of \( w \). This remedy appeared to have the effect of linearly accelerating the convergence process by a factor of approximately \( w \).

Hummel and Rosenfeld\(^6\) defined a geometric framework in which relaxation processes can be conveniently studied. They postulated conditions which any relaxation labelling algorithm must satisfy and showed how a variety of updating rules can be formulated to meet these conditions. In particular the scheme in equation (5) and its accelerated version can be regarded as special cases of an approximate log-linear updating process.

As relaxation processes are essentially parallel, it is easy to envisage that the computation can be done by an array of identical processors, each associated with a single object–label pair in the network. Ulman\(^3\) and Faugeras\(^4\) have investigated the types of problems that such distributed processors can effectively solve and showed that these include relaxation labelling viewed as a constrained optimization problem. They also obtained the conditions under which a relaxation process will be local and discussed the implication of these conditions on object interconnections and network symmetry.

The difficulties of establishing convergence properties of the heuristic updating schemes motivated Faugeras and Berthod\(^40,46\) to formulate the problem of consistent labelling as an optimization problem. They introduced an objective function which comprises two components: a criterion of consistency, and a measure of ambiguity. As in linear relaxation, labelling may be considered consistent when the current probability estimate \( P(\theta = \lambda_i | \lambda_j) \) is equal to the support \( O(\theta = \lambda_i | \lambda_j) \quad \forall i,j \) but other definitions of consistency have been suggested.\(^7\) Ambiguity of labelling, on the other hand, is measured in terms of a quadratic entropy function. The overall objective function is then optimized with respect to label probabilities subject to the constraint that the optimal solution satisfies

\[ \sum_{i=1}^{m} P_{\text{opt}}(\theta = \lambda_i) = 1 \quad \forall i \]

(8)

\[ P_{\text{opt}}(\theta = \lambda_i) > 0 \quad \forall i,j \]

(9)

This new outlook stimulated attempts to view the updating scheme in equation (5) as an algorithm optimizing some objective function which quantifies labelling consistency. This function has eventually been identified by Berthod and Faugeras\(^30\) as having the form

\[ F = \sum_{i=1}^{m} \sum_{j=1}^{n} P(\theta = \lambda_i) O(\theta = \lambda_j) \]

(10)

The form of this function has been adopted by Hummel and Zucker\(^49,51\) as their basic consistency criterion which can be generalized to cater for higher-order compatibilities (triplets, ... \( n \)-tuples of objects).

With the identity of the consistency function established, efforts were made to improve on the optimization procedure embodied in equation (5). Lloyd\(^32\) has rewritten the probability updating formula as

\[ P^{i+1}(\theta = \lambda_i) = P^i(\theta = \lambda_i) + \alpha \rho_0 \]

(11)

where

\[ \rho_0 = \frac{P^i(\theta = \lambda_i) \left[ O^i(\theta = \lambda_j) - \sum_{k=1}^{m} P^i(\theta = \lambda_k) O^i(\theta = \lambda_k) \right]}{\sum_{k=1}^{m} P^i(\theta = \lambda_k) O^i(\theta = \lambda_k)} \]

(12)
She showed that the probability update direction defined by $p_{ij}$ leads to an extremal point of function $F$ and that the nonlinear probabilistic relaxation algorithm given by equation (5) corresponds to setting the stepsize $\alpha$ equal to unity. The algorithm convergence can be considerably accelerated by choosing an optimal value of $\alpha$. Since the objective function $F$ is quadratic, the optimal step can easily be determined.

Following the foundations laid by Hummel and Rosenfeld, Mohamad et al have developed an optimization procedure along more conventional lines. In general the candidate direction of update is defined in terms of the gradient direction of function $F$ at the current label probability values. The actual direction of update is obtained by projecting the gradient vector onto the planes defined by equation (8). In contrast with Lloyd's approach, this optimization process does not terminate when any one of the probabilities assumes the value zero. Even zero probabilities are updated, provided the new values become positive as a result of the update.

If the emphasis is on minimizing ambiguity and the network of objects has an open chain topology, then the probabilistic labelling problem can be solved by means of an iterative application of dynamic programming. This methodology can be applied to suitably structured two-dimensional networks such as lattices by optimizing at each iteration, first along the rows and then along the columns of the network nodes.

Generally, relaxation labelling procedures will find a local extremum of the consistency/ambiguity criterion function employed. As discussed by Faugeras, for example, this should not necessarily be considered a drawback. A local optimum is more likely to reflect the information content introduced by measurements $X$ than a global optimum which will invariably be independent of the input data. However, for some image matching applications which are formulated as consistent labelling problems it may be important to find the global optimum corresponding to the best possible match. In such situations, stochastic optimization as discussed by Geman and Geman, Cohen et al, Flinton and Trivedi, will have to be deployed instead of deterministic procedures.

### Compatibility and Support Functions

As pointed out by several authors (eg Hummel and Rosenfeld, Zucker and Mohamad and Haralick et al), one of the main problems of designing a relaxation labelling process lies in selecting suitable compatibility and support functions. Because of their heuristic basis, the compatibility and support functions in equations (2) and (3) have been found to suffer from a number of drawbacks which include bias and lack of interpretability of the computed probabilities.

Any bias of a relaxation process is best manifested in two simple situations: the independent-node case, and the no-information case. The properties of equation (5) under the assumption of independent nodes were first studied by Pavlidis. When the nodes in a network do not convey any contextual information about each other, the initial label probabilities should not be changed by the updating formula. Pavlidis showed that the algorithm of Rosenfeld et al did not possess this basic property.

Similarly, when all labels at every node have equal initial probabilities, ie the measurements $X$ contain no useful information, an unbiased process should leave these probabilities unchanged. Zucker and Mohamad and Haralick et al derived the conditions which must be satisfied by the compatibility and support functions to eliminate bias both in the general case and in the homogeneous case where all nodes have the same label sets, reciprocal compatibility relationships and equal node weighting. They also showed that all fixed points of the updating scheme in equation (5) can be determined and their stability investigated by means of the Fréchet derivative. A more general characterization of solutions yielded by probabilistic relaxation algorithms is given by Elfving and Eklundh.

The limited justification of the arithmetic averaging for the support function in equation (5) motivated Zucker and Mohamad to consider more general forms for combining the support from the objects conveying contextual information. They showed that the arithmetic average is a special case of the geometric average and advocated the use of the latter.

In the same paper Zucker and Mohamad also introduced a simple transformation which permitted the compatibility coefficients to be viewed as conditional probabilities. The aim of this transformation was to get a better handle on the relaxation process properties and to facilitate its interpretation. Although this simplistic change in itself did not contribute much to the body of knowledge relating to relaxation labelling, it may have inspired others to approach the problem of determining compatibility and support functions in a probabilistic framework.

By setting out to combine the evidence furnished by individual nodes sequentially, Kirby showed that the support function should be of a product form. This result provided a theoretical justification for the heuristic geometric-average rule of Zucker et al.

Other authors were more concerned with compatibility functions. Peleg and Rosenfeld investigated the effectiveness of mutual information as a compatibility function in relaxation labelling. The use of mutual information was also advocated by Yamamoto (referred to by Narayanan). In later work, the use of the somewhat unnatural concept of probability distributions of label probabilities enabled Peleg to apply formal rules for manipulating conditional probabilities and to derive, under fairly restrictive assumptions, an expression for the probability of the joint occurrence of label $i$ at unit $a$ and label $j$ at unit $b$, given the current values of the respective label probabilities. Fortunately some of the most suspect probabilistic entities cancel out when they appear in both the numerator and denominator of the updating formula used. However, the main conclusion of this analysis, namely that the compatibility coefficients should be computed as the ratio of the joint label probabilities and the product of marginal probabilities, has met with little controversy. Most importantly, the relaxation process with compatibility coefficients defined in this fashion can be shown to be unbiased both in the no-information
and the independent node cases. An attempt to establish a relationship between Peleg's approach and the compatibility coefficient of Rosenfeld et al. has been reported by Wang and Zhuang.

In general, the approaches to determining compatibility functions discussed so far can handle higher-order dependences. However, the actual number of studies in which higher-order dependences were experimented with are very few. Some results have been reported by Peleg, Eklundh and Rosenfeld and Fekete et al.

The work on contextual statistical classification algorithms reported by Kittler and Fögelin established close links between the recursive implementation of these algorithms and probabilistic relaxation. This suggested that both compatibility and support functions should be derived using the mathematical apparatus of statistical decision theory. The approach advocated has been illustrated in a number of specific examples by Kittler and Fögelin. The main implication of this work is that compatibility and support functions depend on the assumptions that can be made about the problem in hand. The approach also offers a mechanism for incorporating contextual information in the case of nonmemoryless noise.

An interesting alternative to the above methods of determining compatibility and support functions is advocated by Feugas. He casts the problem of gathering evidence for relaxation labelling in the framework of Schaefer's theory of mathematical evidence. The approach provides a direct mechanism for handling ignorance, i.e., the situation where none of the possible labels can be assigned with reasonable confidence. However, a reject or no-label option can also be introduced in the usual probabilistic framework as discussed by Berthod.

**GENERAL COMMENTS**

The complexity of image understanding processes necessitates that image description be computed at multiple levels of abstraction. Although the methodology of relaxation processes has been developed to deal primarily with the object-labelling problem at any single level of image representation, it can be extended to handle interlevel interaction as demonstrated by Zucker and Mohammad. This potential may prove to be very important, as the ability to draw on information from higher levels of description which is of a more global character is considered essential for maximum reduction of ambiguity and hypothesis verification. A similar approach has been pursued by Kuschel, Kuschel and Page and Davis and Rosenfeld. Glazer and Narayanan et al. recommend a multilevel relaxation process involving either the raw image or one of its lower level of abstraction at different resolutions. The main aim of the multiresolution representation is to increase the rate of convergence of the relaxation process.

A frequently observed behaviour of relaxation labelling processes is an initial reduction in ambiguity of interpretation which peaks after the first few iterations and is then followed by gradual degradation of the labelling performance. Haralick set out to explain this characteristic by attempting to provide interpretation for the probabilities yielded by the updating process. He achieved some degree of success with the product updating rule where the compatibility coefficients were defined as by Peleg and also argued that the conceptual framework in which Peleg's relaxation algorithm was developed precluded any meaningful interpretation of the computed probabilities. This argument, however, is largely academic as Peleg's relaxation labelling scheme involving the product rule is identical to that interpreted by Haralick. While the relaxation process involving the arithmetic-average rule is less obvious, nevertheless the interpretation for the probabilities computed by the classical scheme put forward by Kittler and Fögelin is directly applicable also to Peleg's updating formula and renders it less objectionable.

The need for interpretation is less important when compatibility and support functions are derived as advocated by Kittler and Fögelin as the assumptions underlying updating processes are stated a priori and do not have to be deduced retrospectively. In general, however, even here the interpretation is valid only for the first iteration. The meaning of the probabilities yielded by subsequent iterations is increasingly speculative.

To obviate the lack of adequate theory which would allow the selection of well-behaved compatibility and support functions, Peleg and Rosenfeld adopt a pragmatic view and devise evaluation measures which can be used as criteria for stopping the relaxation process. They suggest that an evaluation measure should reflect the best compromise between the information conveyed by measurements and the contextual information conveyed by the model of the relationships between objects. The problem has also been addressed by Elfving and Eklundh. Richards et al. on the other hand suggest heuristic rules that must be satisfied by the compatibility and support function so that the performance deterioration is overcome. These rules can be used as the basis for a possible probabilistic relaxation procedure design strategy. As an alternative approach Richards et al. advocate the use of a modified updating scheme where the label probabilities computed according to equation (5) are adjusted by a factor which is dependent on the initial label probabilities.

With the increasing proliferation and variety of relaxation labelling techniques, a need has been identified to carry out comparative studies to assess the relative advantages and drawbacks of individual methods. However, a comprehensive systematic comparison of many methods on a range of data represents a large research effort and as yet few studies have been reported. O'Leary and Peleg consider the effect of several relaxation methods in the simple case of labelling two nodes with a set of two labels, while Nagin et al. have performed experiments on segmenting test images which contain a variety of spatial structures using a range of compatibility coefficients and/or neighbourhood configurations. At a higher level of interpretation, Price has given some results for a number of methods applied to the relaxation matching of semantic network descriptions of images but his
work is only briefly reported and a more detailed de-
scription of methods and results should be compiled.
Some comparative results have also been reported by
Fekete et al.\textsuperscript{22}

Although these studies throw new light on a number
of aspects of various methods, without exception they
have proved inconclusive as they tend to be unsys-
tematic and compare different updating schemes that
have different compatibility and support functions.

APPLICATIONS

Toy problems

Several authors have considered relaxation on particu-
larly simple graph structures with only a few labels.
The most famous example is the three-node-four-label
toy triangle example introduced in the original paper
by Rosenfeld et al.\textsuperscript{49}

However, as mentioned above, O’Leary and Peleg\textsuperscript{96}
have studied a two-node-two-label graph and
Haddow\textsuperscript{80} has analysed the relaxation scheme
of Rosenfeld et al.\textsuperscript{49} for the particularly symmetric case of
a tetrahedral graph with a two-label set at each node.
All these are nontrivial examples which produce a range
of solutions dependent on both the initial probabilities
and the assumed compatibilities between labels on
pairs of connected nodes. They are particularly useful
in helping to understand the nature of fixed points
of the relaxation process.

Classification

Bhanu and Faugeras\textsuperscript{60} have used relaxation to binarize
a grey-level image. Initial probabilities $p_a$ and $p_b$ of
the pixel being object or background are assigned on
the basis of individual pixel grey level and these are
then iteratively updated using the supporting evidence
$q_a$ and $q_b$ for the labels from the immediate-neighbour
pixels. They use a criterion function approach which
is based on the dot product of the pixel probability
vector with its neighbourhood support vector. By maxi-
mizing the criterion for the sum of all the pixels, they
hope to obtain the most consistent and least ambigu-
ous label assignments. Optimization is achieved by
a projected gradient method which finds the nearest
local maximum to the initial labelling. Good results,
which can be controlled by variation of a few par-
eters pertaining to the initial assignment scheme and
the optimization procedure, can be obtained.

The use of relaxation for the more general case of
improving many-label classification of multispectral
data has been attempted by many authors, especially
in relation to remotely sensed data\textsuperscript{81,82,83,84} and
colour images\textsuperscript{77,95,96}. Most of these applications have
been approached in the same way. Model clusters are
defined in the measurement hyperspace either by auto-
matic clustering or by hand segmentation of ‘ground
truth’ data. The initial label probabilities are cal-
culated as a simple function of the Mahalanobis distance
between the models and the measured pixel data. The
interrelationships of labels are derived empirically from
‘ground truth’ data by measuring and globally averag-
ing transitional probabilities, correlations or the mutual
information\textsuperscript{86,89} of neighbouring pixels. Upon applying
these methods, most investigators report a sharp initial
decrease in classification error rates of several per cent
followed by a smaller increase as the process converges
to a stable solution. Kalayeh and Landgebre\textsuperscript{97} suggest
that many problems can be overcome by deriving esti-
mates of compatibility coefficients over a local, say
5 × 5 pixel, window.

Edge enhancement

Labelling points as edge pixels is another important
application where relaxation has been tried\textsuperscript{98}. Schacter
et al.\textsuperscript{50} posed this problem as an edge/no-edge classi-
ﬁcation problem and used edge direction information
to provide measures of compatibility between neigh-
bouring pixels. Edge label probabilities were adjusted
using a heuristic function, and at the end of each iter-
ation the edge directions were reestimated as
weighted averages of edge directions of all pixels in
a small neighbourhood. The weighting factors were
the probabilities of the pixels being labelled as edges.
Schacter et al. claimed the method proved highly effec-
tive as a noise-cleaning technique but it requires the
adjustment of several arbitrary parameters.

Hanson and Riseman\textsuperscript{96} suggested an edge relaxation
scheme which is based on applying collinearity con-
straints to binary arrays of edge segments. The assign-
ment of the strength of an edge segment was adjusted
on the basis of the number of edge segments meeting
at four-way vertices after thresholding the edge
strength. Between zero and three other edges can meet
an edge at a vertex. If an edge segment occurs which
has no neighbours at either end, then it is completely
isolated and is very unlikely to be part of a true edge.
The edge strength of such a segment is severely
reduced. However, if exactly one other edge occurs
at each end of an edge, then there is good continuation
and the strength assignment of such edges is strongly
supported. Cases where more than two edges meet
at a vertex are treated as ambiguous evidence and edge
label strengths are left unadjusted. Thresholding and
degree strength reassignment are iterated many times.
Results were reported as very favourable.

Faugeras and Bertho\textsuperscript{46} illustrate their optimization
technique for relaxation by considering labelling pixels
with a set of four edge labels (north, east, south, west)
and a no-edge label. Initial edge probabilities were
derived using the normalized output of a standard
differencing 3 × 3 edge operator. Only edge magnitude
was used in the calculation. The criterion which was
optimized was a combination of a consistency measure
(the agreement of label assignment based on evidence
of the pixel and the evidence of its neighbourhoood)
and an ambiguity measure (the quadratic entropy of
the labelling). The compatibility coefficients were
derived empirically from hand-labelled data. A few
iterations of the process produced good results.

Wambacq et al.\textsuperscript{100} have used an edge enhancement
process with a label set similar to that of Faugeras
but using the standard updating scheme of Rosenfeld et al. The results of edge enhancement enable decisions relating to the choice of a suitable post processing image filter to be made more confidently; however, this study also stresses that a suitable methodology for the assignment of initial probabilities, compatibilities and the choice of relaxation updating scheme is lacking.

Danker and Rosenfeld have reported combining edge and region information in a joint relaxation scheme to extract 'blobs' from images. They found that the two processes complemented one another, leading to the suppression of spurious edges within the background image area and prevention of the expansion of blob assignments into the nonblob areas. They also reported that the speed of convergence was improved.

Feature enhancement

Detection of lines, strips and curves has been attempted by several authors. Zucker et al. use a thin rectangular neighbourhood mask oriented in one of eight directions defined by a line operator whose output forms a measurement vector for initial probability assignments. By weighting pixels in the context conveying neighbourhood, the curvature of the lines which are to be enhanced can be controlled. The main difficulty reported is a tendency to produce curves which are too thick. However, a thinning constraint can be incorporated into the compatibility coefficients to overcome this tendency.

Peleg and Rosenfeld have considered how compatibility coefficients can be computed from suitable test data. They found that curves can be found equally satisfactorily using either correlation coefficients or mutual information measures and the coefficients derived from one image were suitable for use in curve relaxation on other images.

Danker and Rosenfeld suggested that strips, which are formed from two antiparallel edges, are important features in many images. They may represent long straight roads, runways or structures such as tree trunks. Relaxation processes can be used to reinforce collinear but antiparallel edge responses and hence to identify strips. Danker and Rosenfeld use a simple iterated convolution method to enhance strips of an appropriate size and orientation.

Shape and stereo matching

Relaxation labelling has been much used in matching problems such as two-dimensional shape matching or the stereo correspondence problem. Representative feature points such as corners are extracted from the template shape and the real-world image. Initial probabilities can be assigned on the basis of the degree of match of these local features and these probabilities can then be iteratively reinforced on the basis of the occurrence of matches with other features which are found in approximately correct relative positions.

The stereo correspondence problem is very similar.

Features are extracted in two images. The features of one image can be regarded as nodes of a graph while the features in the second image are possible labels for the nodes. The initial probabilities of labels can be taken as a simple function of the distance between node and label points in the two images. Node and label assignments are then compatible if similar neighbourhood states exist in both images. Although these relaxation schemes might not have great computational advantages relative to other standard matching methods, they are more tolerant of image distortion. In addition, the reinforcing processes are local and therefore missing matches can be tolerated and hence occluded objects can be recognized.

High-level interpretation

At the highest image interpretation levels, the matching of relational networks has been studied by Kitchen and Rosenfeld. Discrete relaxation is applied to relational networks consisting of areas from maps, atoms of B-codeine and components in a complex switching network. The algorithm converged very quickly in all these cases to a fairly unambiguous result. Kitchen also considered relational networks with numeric as well as symbolic attributes using a fuzzy relaxation matching scheme. Although these methods were sensitive to the exact form of the updating rule, it was found they could work very well and tolerate quite large measurement errors. A similar fuzzy relaxation scheme has been used by Barrow and Tenenbaum in their reasoning system for image analysis, MSYS.

Faugeras and Price used probabilistic relaxation to match high-level semantic network descriptions of images. Images were segmented using an automatic system based on region growing and line extraction processes. Features of the segmented areas such as average colour, average texture, size, shape, position and orientation were derived and compatibility relationships were defined in terms of adjacencies and relative distances and orientations of segments. In a typical image, between 100 and 200 segments might have had to be matched. A world model was defined with similar attributes and a global criterion for the relaxation matching of segments was optimized. The scheme incorporated a probability threshold and if a segment label probability exceeded it, the segment was given this label with certainty. This meant that good matches were quickly made. Faugeras and Price have reported good results, most of the disagreements being attributable to poor initial segmentations.

Other applications and hardware implementations

In addition to applications of direct relevance to images, the relaxation labelling technique has been applied to problems in character recognition, handwriting interpretation, the breaking of substitution ciphers, the reconstruction of matrices from summed projections, the segmentation of curves and wave-
CONCLUSIONS

This survey shows that many significant advances have been made since Rosenfeld et al first introduced probabilistic relaxation labelling. It has shown success or great promise in many application areas. However, many issues still remain to be resolved, especially in respect of defining suitable compatibility relationships, interpreting in detail the effects of some of the suggested updating schemes and proving convergence properties. Relaxation labelling should remain a fertile research field for several years to come.

REFERENCES

1 Ullmann, J R 'A consistency technique for pattern association' Trans. IRE (IT) Vol 8 No 5 (1962) pp 74–81
4 Zucker, S W 'Relaxation labeling and the reduction of local ambiguities' Proc. 3rd Int. Joint Conf. on Pattern Recognition, Coronado, CA, USA (1976) pp 862–867
7 Waltz, D L 'Understanding line drawings of scenes with shadows' in Winston, P H (ed.) The psychology of computer vision McGraw-Hill, New York, USA (1967)
17 Shapiro, L G 'Solving consistent labeling problems having the separation property' Proc. 7th Int. Conf. on Pattern Recognition, Montreal, Canada (1984) pp 313–315
24 Faugeras, O D 'Optimisation technique in image analysis' Proc. 4th Int. Conf. on Analysis and Optimisation of Systems, Tokyo, Japan (1980) pp 790–823
31 Sakaue, K and Takagi, M ‘Image segmentation by iterative method’ Proc. 6th Int. Conf. on Pattern Recognition, Munich, FRG (1982) pp 192–194
45 Faugeras, O D and Berthod, M ‘Scene labelling: an optimisation approach’ Proc. IEEE Conf. PRIP, Chicago, IL, USA (1979) pp 318–326
47 Berthod, M ‘Definition of a consistent labeling as a global extremum’ Proc. 6th Int. Conf. on Pattern Recognition, Munich, FRG (1982) pp 393–401

image and vision computing

55 Diamond, M D, Narasimhamurthi, N and Ganapathy, S ‘A systematic approach to continuous graph labeling with application to computer vision’ Proc. Nat. Conf. on Artificial Intelligence (AAAI-82), Pittsburgh, PA, USA (1982) pp 50–54
57 Cohen, F S, Cooper, D B, Silverman, J F and Hinkle, E B ‘Simple parallel hierarchical and relaxation algorithms for segmenting textured images based on noncausal Markovian random field models’ Proc. 7th Int. Conf. on Pattern Recognition, Montreal, Canada (1984) pp 1104–1107
58 Hinton, G ‘Relaxation: its role in vision’ PhD Thesis Edinburgh University, UK (1978)
59 Trivedi, H P ‘Solving the binocular stereo correspondence problem using simulated annealing’ Proc. 4th Scandinavian Conf. on Image Processing Trondheim, Norway (1985) pp 167–173


70. Peleg, S 'Monitoring relaxation algorithms using labeling evidence' *Proc. 6th Int. Conf. on Pattern Recognition, Miami, FL, USA* (1980) pp 54-57


74. Faugeras, O D 'Relaxation labelling and evidence gathering' *Proc. 8th Int. Conf. on Pattern Recognition, Munich, FRG* (1982) pp 405-412


81. Richards, J A, Landgrebe, D A and Swain, P H 'Overcoming accuracy deterioration in pixel relaxation labeling' *Proc. 5th Int. Conf. on Pattern Recognition, Miami, FL, USA* (1980) pp 61-66


88. Price, K E 'Relaxation matching techniques — a comparison' *Proc. 7th Int. Conf. on Pattern Recognition, Montreal, Canada* (1984) pp 987-989


Lloyd, S A 'A dynamic programming algorithm for binocular stereo vision' GEC J. Res. Vol 3 No 1 pp 18–24

Lloyd, S A 'A dynamic programming algorithm for binocular stereo vision' GEC J. Res. Vol 3 No 1 pp 18–24

Marr, D and Poggio, T 'Cooperative computation of stereo disparity' Science Vol 194 (1976) pp 283–287


Barrow, H G and Tenenbaum 'MSYS: a system for reasoning about scenes' Tech. Note 121 Stanford Research Institute, Menlo Park, CA, USA (April 1976)


Velasco, F R D and Rosenfeld, A 'The application of relaxation to waveforms with ambiguous segmentations' IEEE Trans. SMC Vol 9 No 8 (1979) pp 420–428


McLean, C R and Dyer, C R 'An analog relaxation processor' Proc. 5th Int. Conf. on Pattern Recognition, Miami, FL, USA (1980) pp 58–60