

CONTINUOUS RELAXATION AND LOCAL MAXIMA SELECTION

Conditions for Equivalence**

Steven W. Zucker, Yvan G. Leclerc, John L. Mohammed
Computer Vision and Graphics Laboratory
Department of Electrical Engineering
McGill University
Montreal, Quebec, Canada

It is also a good rule not to put overmuch confidence in the observational results that are put forward until they are confirmed by theory.

Sir Arthur Eddington, 1934

1. MOTIVATION

The study of speech and image understanding systems has largely been empirical. Representations have often been designed for task domains, while algorithms were designed to make use of these representations, as well as heuristic considerations about how the algorithm should behave. Thus it has been very difficult to compare different algorithms except for the results of a few examples, to state equivalences between algorithms in anything like certain terms, or to determine the theoretical requirements for solving a particular problem.

A number of common problems facing the designers of such systems have begun to emerge, however, which suggests that it may be possible to study the power of algorithms in terms of their ability to solve these problems. One such problem is the reduction of local ambiguities, and relaxation labeling processes are a class of algorithms that have been proposed for dealing with it [1]. While many recent applications offer some insight into the relaxation computation (for a review, see [2]) a formal analysis aimed at specifying this computation more precisely is required. It is impossible, for example, to know whether two different algorithms are really equivalent (in the sense that they are implementing the same abstract computation) simply by comparing their results on a small number of experiments. Rather, the relationships between them must be expressed in an analytical fashion so that they will hold for all experiments. This paper is the beginning of such an analysis for continuous relaxation processes. It specifies the relaxation computation in terms of another, much more widely used algorithm - local maxima selection - together with a set of conditions that are sufficient for guaranteeing convergence and

equivalence between the two.

The specific reasons for studying the relationships between relaxation and local maxima selection are manifold. From the relaxation point of view, maxima selection provides a model for the decision process, or part of the decision process, that relaxation is implementing. In other words, if we consider the dynamics of a relaxation process as being composed of two stages, one in which a proper ordering of the labels is obtained given the compatibility relations and the initial state, and the other in which certain of these labels are selected, then maxima selection provides a non-iterative model for the second stage. From a practical point of view, when conditions sufficient for equivalence between the two algorithms are met, they can serve as stopping criteria for the relaxation process. More generally, however, these results show that when circumstances are structured well enough, the cooperative relaxation algorithm acts the same as if it were decomposed into a purely local maxima selection process. This is exactly the kind of relationship that is desirable between two algorithms: when the stronger one is necessary, it is functional; otherwise, it reduces to the weaker one.

There is another, more abstract application of the theory that we are beginning to develop here. It derives from the decomposition result just alluded to (which is stated more formally in Sec. 3), and pertains to inferences about how local or global a given algorithm must be. In

*This research was supported by NRC grant No. A4470.

+For a much more extensive discussion for these ideas, together with references, see TR-78-15R, Dept. of Elect. Eng., McGill University.

particular, cooperation can be viewed as one way in which information can be propagated between the local parts of an algorithm, hence making it more global [3]. One surprising interpretation of our result is that circumstances exist in which this propagation is irrelevant. In other words, although a (globally) cooperative mechanism has been used in designing an algorithm for a given problem, the inherent structural complexity of the problem (in that representation) is still only local. Conditions such as those stated in the Equivalence Theorem (Sec. 3) provide a first step toward determining such requirements analytically. In turn, these computational requirements may imply certain constraints on the particular mechanisms and machinery that actually perform such functions [4].

2. BACKGROUND

The class of problems for which both relaxation and local maxima selection are applicable is one in which global labeling problems can be represented as networks of local ones. In other words, given a set of possible labels for each node in a network, the problem is to select a single label (or a smaller set of labels) to attach to each node. These labels might indicate, e.g., assertions about the presence of oriented line segments in an image [5]. The nodes in the network would then correspond to the spatial positions to be labeled, while edges would indicate which nodes were spatial neighbours. Or, the labels might indicate disparity values [6]. We shall denote such labels by the variable λ , and the full set of labels (at a node i) by Λ_i .

For continuous relaxation processes, a certainty measure is distributed over the label set at each node. We shall denote the certainty with which label λ is associated with node i by $p_i(\lambda)$. They satisfy $\sum_{\lambda \in \Lambda_i} p_i(\lambda) = 1$ and

$0 \leq p_i(\lambda) \leq 1$. In the line labeling example, such a measure may be obtained from scaled feature (i.e., line) detector responses. Note that ambiguity enters the problem because feature detectors do not respond uniquely; i.e., many labels appear to be possible when looking only at the individual detector responses.

Perhaps the most common algorithm for selecting a label from among this initially ambiguous set is to perform local maxima selection (i.e., to pick the label with the maximal certainty at each node). Such an algorithm would be appropriate for our line example, if we knew

that the line detector always responded most strongly at the correct orientation. This is, of course, not the case for real images, in which many different noise sources can corrupt the detectors' responses. Thus more powerful techniques seem to be necessary in general. However, as we shall show, there are instances in which such weaker techniques are appropriate.

The strength behind relaxation processes lies in their use of contextual information throughout the ambiguity reduction process. This context is based on an a priori notion of consistency or compatibility between labels on neighboring nodes, so that now some degree of cooperation takes place between nodes during the label selection process. In the line labeling example, label compatibility is an abstract way to introduce models for good continuation of orientation (i.e., orientation should vary smoothly along lines and curves) and of intensity (i.e., intensity values should vary smoothly along the line and over the background); see [7]. In this paper, however, since we are discussing the relaxation mechanism in the abstract, we shall only consider compatibilities as functions over n -tuples (usually pairs) of labels on neighboring nodes into $[0,1]$. For example, $r_{ij}(\lambda, \lambda')$ indicates the compatibility between label λ on i and λ' on its neighbor j , where 0 corresponds to perfect incompatibility and 1 to perfect compatibility.

Given an initial labeling $p_i^0(\lambda)$, $\lambda \in \Lambda_i$, $i \in I$, the relaxation process iteratively updates each label's certainty factor by an amount proportional to an estimate of its consistency with the labeling over its neighborhood. If this neighbourhood consistency estimate indicates that a label is compatible with its neighbourhood labelings, then its consistency factor is increased; otherwise, it is decreased. Since many labelings may be partially consistent, the update is normalized with respect to the average neighbourhood consistency at each iteration. Thus we have updating rules of the form:

$$p_i^{k+1}(\lambda) = p_i^k(\lambda) \cdot \frac{Q_i^k}{\sum_{\eta \in \Lambda_i} p_i^k(\eta) Q_i^k(\eta)} \quad (2)$$

where $Q_i^k(\lambda)$ is the current consistency estimate between label λ on node i and the labeling over its neighbourhood.

The neighbourhood consistency is computed as a function of the current neighbourhood labeling and the compatibility functions. For comput-

ational reasons, this estimate has been decomposed into two parts: one in which an operation is performed to determine the pair (or n-ary) label consistency estimates; and the other in which these partial estimates are combined into a single neighbourhood estimate. The most widely used estimate is [1]:

$$Q_i^k(\lambda) = \sum_{j \in J_i} C_{ij} \sum_{\lambda' \in \Lambda_j} r_{ij}(\lambda, \lambda') p_j^k(\lambda') \quad (3)$$

where J_i is the set of nodes neighbouring i and the C_{ij} terms weight the total interaction between nodes i and j . The choice of arithmetic averages in (3), however, is only heuristic. The essential point is that the neighbourhood consistency estimate is a function of the labeling over the entire neighbourhood. To emphasize this generality, we shall rewrite (3) in a more general form, and prove all of our results using this form.

The general form that we shall consider is:

$$Q_i^k(\lambda) = \text{Mean} \{ \text{Norm} r_{ij}(\lambda, \lambda') p_j^k(\lambda') \} \quad (4)$$

where Mean is any standard mean (or averaging) operator; and Norm is an operator that computes the norm (or length) of a vector.

3 CONSISTENTLY ORDERED LABELINGS AND THE EQUIVALENCE THEOREM

The traditional way of viewing continuous relaxation is as numerical process for updating certainty measures. However, if we consider these certainty measures as defining a partial ordering over the labels at each node, then another way of viewing relaxation labeling is as a process that changes this partial ordering. This is the point of view that we shall adopt for the remainder of this paper, because it suggests a study of the iteration at which the partial ordering of the labels at each node becomes fixed for all further iterations. We shall say, at this iteration, that the system has become consistently ordered. The particular partial ordering that we shall concentrate on is the one specifying the relationship between the maximal and all of the non-maximal labels at each node.

There are many possible dynamical situations that can give rise to consistently-ordered labelings. Perhaps the most common situation is for the maximal labels to mutually support one another, thus driving the weaker labels to zero. It is this situation that we shall now

attempt to characterize because, as we shall prove, it leads to a formal correspondence between relaxation and local maxima selection.

To define the above notions more exactly, it is necessary to introduce some notation: Let

- (i) Λ_i be the full set of labels attached to node i .
- (ii) M_i^k be the set of labels attached to node i that have the greatest certainty at iteration k ;
- (iii) N_i^k be the set of labels attached to node i with non-zero, non-maximal certainties at iteration k ;
- (iv) Z_i^k be the set of labels attached to node i with zero certainty at iteration k .

Equivalence Theorem

If, for a given relaxation process, the conditions

- (i) $\text{Norm} r_{ij}(m_i, m_j) p_j^{k_C}(m_j) > \text{Norm} r_{ij}(n_i, \lambda') p_j^{k_C}(\lambda')$
 $m_j \in M_j^{k_C}, \lambda' \in \Lambda_j$

$$\forall m_i \in M_i^k, \forall n_i \in N_i^{k_C},$$

$$\forall j \in J_i$$

and

- (ii) $r_{ij}(m_i, \lambda') = r_{ij}(m_i', \lambda')$

$$\begin{aligned} & \forall m_i, m_i' \in M_i^{k_C}, \\ & \forall \lambda' \in M_j^{k_C} \cup N_j^{k_C}, \\ & \forall j \in J_i \end{aligned}$$

hold for every node $i \in I$, then

- (a) the labeling $\mathcal{L}^{k_C}$ in that process is consistently ordered everywhere, and
- (b) the relaxation process will converge to a final labeling with

$$\lim_{k \rightarrow \infty} p_i^k(\lambda) = \begin{cases} 1/\#M_i^{k_O} & \forall \lambda \in M_i^{k_O} \\ 0 & \text{otherwise} \end{cases}$$

where $\#M_i^{k_O}$ denotes the number of labels in $M_i^{k_O}$. For the proof of this theorem, together with applications and interpretations, see the TR cited in the footnote on the title page.