

*The least-disturbance principle and weak constraints**

Andrew BLAKE

Machine Intelligence Research Unit, Edinburgh University, Edinburgh, Scotland

Received 3 May 1983

Abstract: Certain problems, notably in computer vision, involve adjusting a set of real-valued labels to satisfy certain constraints. They can be formulated as optimisation problems, using the 'least-disturbance' principle: the minimal alteration is made to the labels that will achieve a consistent labelling. Under certain linear constraints, the solution can be achieved iteratively and in parallel, by hill-climbing. However, where 'weak' constraints are imposed on the labels – constraints that may be broken at a cost – the optimisation problem becomes non-convex; a continuous search for the solution is no longer satisfactory. A strategy is proposed for this case, by construction of convex envelopes and by the use of 'graduated' non-convexity.

Key words: Constrained labelling, relaxation, optimisation, non-convexity, parallel array processor, computer vision.

1. Introduction

Constrained labelling can be described formally as the problem of affixing labels to the nodes of a graph, subject to constraints which act along arcs between nodes. This is achieved by combining information in an initial labelling or set of candidate labellings, with information in the constraints, to generate a new labelling. The labels themselves may be discrete (Waltz 1972) or continuous (Ikeuchi 1980, Rosenfeld et al. 1976). In each case, the literature reports the use of parallel iterative algorithms to find a consistent labelling. For applications in low-level image processing, nodes of the graph often correspond to the pixels of an image array – one or more nodes per pixel – and constraints act along arcs connecting each node to nodes on some of its neighbouring pixels. Parallel, iterative algorithms are attractive here, because they can use the power of parallel array processors

like CLIP4 (Duff 1978). This paper addresses problems in which an initial estimate of the labelling is available. This labelling is to be adjusted to comply with constraints by relaxation. For example, in Hanson and Riseman's edge relaxation scheme (Hanson and Riseman 1978) the idea is that edge segments in an image, lying between pairs of adjacent pixels, are labelled with strengths subject to continuity constraints: edge segments should preferably form continuous chains, representing object boundaries in the image. In successive iterations, the support for the continued existence of an edge element is gauged from the strengths of neighbouring edge elements, and its strength is enhanced or diminished accordingly.

Algorithms described in the literature for adjusting real-valued labellings by relaxation have used non-linear formulae to calculate the successive adjustments (Hanson and Riseman 1978, Marr 1978, Rosenfeld et al. 1976). The effects of the formulae are not fully understood, although some analysis has been done (Harrick et al. 1980, Zucker et al. 1981). In particular, it is not clear what computation these non-linear algorithms perform – the labelling produced does not necessarily

* This research was conducted with the aid of a grant from the SERC to Professor Donald Michie for work on a versatile programmable industrial image processor. The author is also indebted to the University of Edinburgh for the provision of facilities.

satisfy any particular set of constraints. Nor is it guaranteed that the relaxation processes will converge, although they appear to do so in some cases.

There are instances, however, of other relaxation schemes based on optimisation, whose computational effect and convergence properties are well defined (Hinton 1978, Ikeuchi 1980, Ullman 1979). This paper proposes the use of the 'least-disturbance' principle with real-valued labels, for problems in which an initial labelling is to be adjusted so that it satisfies a given set of constraints. The initial labelling is coaxed into satisfying the constraints by the most direct available route; optimisation is used to obtain the consistent labelling that is nearest to the initial labelling, measured by Euclidean distance (though other distance measures may be appropriate in some cases). In this way, information in the initial labelling is preserved, as far as possible, for subsequent processes. Under certain linear constraints, the optimisation problem can be solved by a convergent, linear, iterative method. This approach has been used for edge relaxation and the resulting algorithm implemented on an emulator of a parallel array processor (Blake 1983).

This paper examines some labelling problems in computer vision in which the constraint on the labels is that they are continuous 'almost everywhere'. This is expressed formally as a 'weak constraint' which produces, however, a mathematically non-convex optimisation problem. A continuous search for an optimum is no longer adequate because any of a number of local optima may be found, of which only one is the desired globally optimal solution. A strategy for dealing with those problems is discussed.

2. Weak constraints and non-convexity

Strokes (oriented line segments) have been used as low-level, primitive features in an image (Perkins 1978) and the idea of detecting straight lines of strokes by non-linear relaxation is discussed in the literature (Zucker et al. 1977). Can this task be expressed using explicit collinearity constraints, and the least-disturbance principle?

Consider a chain of N strokes, each labelled

with an angle θ_i , $i=1, \dots, N$, initially $\theta_i^{(0)}$. It may be possible, by adjustments to the angles, to fit the strokes to a piecewise linear model (Figure 1). The adjustment is acceptable if it is not too extensive. Otherwise part or all of the chain is deemed a poor fit to the model. Intuitively, the piecewise collinearity constraint is that strokes are collinear 'almost everywhere'. This idea can be represented more precisely as a constraint that may be broked at a cost – what Hinton (1978) calls a weak constraint. Each pair of adjacent strokes must either be collinear ($\theta_i = \theta_{i+1}$), or else the junction of the two strokes is labelled as 'discontinuous' and a fixed penalty, c^2 , is incurred. The least-disturbance labelling under these constraints is obtained by minimising an objective function:

$$F = \sum_i (\theta_i - \theta_i^{(0)})^2 + c^2 \times (\text{number of broken constraints}), \quad (1)$$

a distance measure on $\bar{\theta} - \bar{\theta}^{(0)}$ plus a penalty of c^2 for each broken collinearity constraint.

The idea of constraints which hold almost everywhere is an important one. For instance, applied to an intensity image it yields a segmentation of the image. Discontinuities that remain after the constraint has been imposed mark region boundaries and other features that can then be used as the basis of higher-level analysis. The penalty constant c gives control over the reluctance to allow discontinuities. A high value of c results in a broad brush segmentation marking only the principal discontinuities, whereas a smaller value results in the inclusion of more detail. Marr (1978) also uses weak continuity explicitly in his cooperative stereo algorithm: binocular disparity is continuous almost everywhere. He uses a non-linear relaxation algorithm, in which it is not entirely clear when

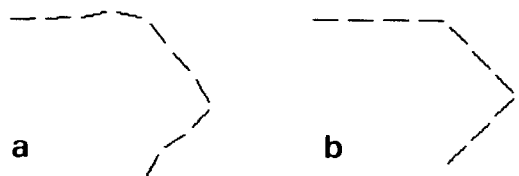


Fig. 1. A chain of strokes (a), organised into straight lines (b), by making small adjustments to their angles.

continuity is enforced and when it is waived. The weak constraint as described above, incurring a penalty when it is broken, attempts to make this more precise.

Unfortunately, the optimisation problem that the weak constraint generates has the mathematical property of non-convexity. This means that, unlike problems in which the constraints are linear (e.g. Blake 1983), there may be a number of local optima so that a continuous search for a local optimum does not guarantee to locate the required global optimum. To illustrate, consider again arranging strokes into straight lines. Figure 2 shows three different arrangements of a chain of strokes. Starting with Figure 2a, continuously and simultaneously adjusting the stroke angles to reach Figure 2c, the chain passes through intermediate states like Figure 2b. In Figure 2b, the collinearity constraint is broken once, incurring a penalty, c^2 , while in Figure 2c the strokes are collinear – no penalty. Let us assume that in Figure 2c, the remission of the penalty outweighs the increase in the distance measure in moving from Figure 2a to 2c, so that Figure 2c is a more desirable state than Figure 2a. Then clearly, Figure 2b bears both a penalty and an increased distance measure, and is therefore less desirable than either of Figure 2a or 2c. Figure 2c cannot be reached by following a strategy of continuous improvement and this is precisely what is meant by 'non-convexity'.

If a strategy of continuous improvement is inadequate, what are the alternatives? Berthod (1982) and Ullman (1978) both suggest settling for a local optimum as the best available solution to a non-convex problem. This is quite inapplicable here, however, because the initial labelling is already a local optimum of F in (1). Hill-climbing would make absolutely no change to the initial state. Another possibility is to use a search, in which numerous configurations of the chain of strokes are tried, but the number of configurations

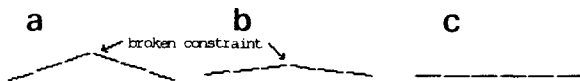


Fig. 2. An initial set of strokes (a) is adjusted continuously (b) until it is collinear (c).

to examine is related exponentially to the number of strokes in the chain. The search is impractical unless it can be guided in some way. This paper proposes instead the use of convex envelopes to replace the objective function F with some other function whose global minimum is the same, but which is convex.

For example, Figure 3a shows the objective function, f , for the chain in Figure 2, simplified by forcing all the strokes to move together; there is only one variable – the deviation, θ , of each stroke from its initial position. The function is non-convex, but we can construct a 'convex envelope' – a convex function, f^* , such that everywhere $f^*(\theta) \leq f(\theta)$ and with the same global minimum (Figure 3b). Continuous descent over f^* follows a steadily improving path towards the global minimum, where all the strokes are collinear.

In this case, where F is f , a function of one variable only, construction of a convex envelope is straightforward. Generally when F is a function of many variables it may be much harder. However, in labelling problems in which the constraints are local, F can be decomposed as the sum of a number of local functions. For instance in the stroke labelling problem we have, from (1),

$$F = \sum_i (\theta_i - \theta_i^{(0)})^2 + c^2 \sum_{i=1}^{n-1} [\theta_i \neq \theta_{i+1}], \quad (2)$$

where $[]$ is used to mean

$$[b] = \begin{cases} 0 & \text{if } b \text{ is false,} \\ 1 & \text{if } b \text{ is true.} \end{cases}$$

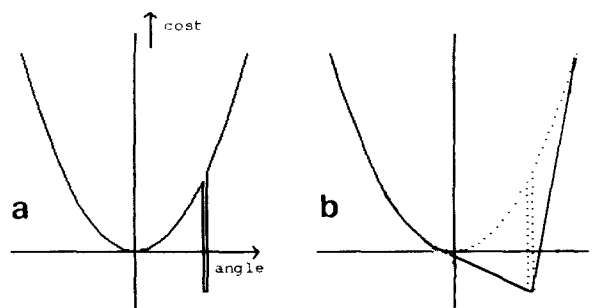


Fig. 3. (a) The cost function for Figure 2. The sharp dip occurs when the strokes are collinear, and makes the function non-convex. (b) A convex envelope of the cost function.

The expression for F can be rearranged to give:

$$\begin{aligned} F(\theta_1, \dots, \theta_N) = & (\theta_i - \theta_i^{(0)})^2 \\ & + \frac{1}{2}(\theta_1 + \theta_2 - \theta_1^{(0)} - \theta_2^{(0)})^2 \\ & + \dots + \frac{1}{2}(\theta_{n-1} + \theta_n - \theta_{n-1}^{(0)} - \theta_n^{(0)})^2 \\ & + (\theta_n - \theta_n^{(0)})^2 + \sum_{i=1}^{n-1} h_i(\theta_i - \theta_{i+1}). \end{aligned} \quad (3)$$

The non-convex part of F is the last term and each of the functions h_i looks something like Figure 3a. Now we replace each h_i by its convex envelope (Figure 3b) h_i^* , to produce a function F^* , which is itself convex (a sum of convex functions is convex). However, it is not in general a convex envelope of F . We can find the global minimum of F^* , because it is convex, but this may or may not be the global minimum of F . However, it is a straightforward task to determine whether it is or not: having found the global minimum, $\bar{\theta}^*$, of F^* , it can easily be shown that

$$F^*(\bar{\theta}^*) = F(\bar{\theta}^*) \quad (4)$$

is a necessary and sufficient condition that $\bar{\theta}^*$ is the global minimum of F .

3. Graduated non-convexity

When the convex envelope method fails to produce the global optimum (and this will be apparent because (4) is not satisfied) a strategy is required to obtain a good sub-optimal solution. A good

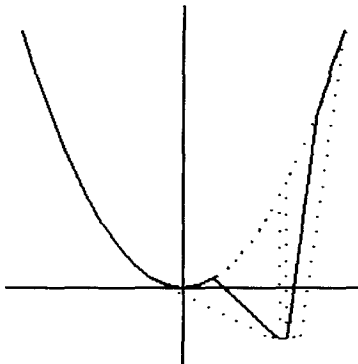


Fig. 4. An intermediate function, lying between the cost function and its convex envelope. A family of such functions is used in the graduated non-convexity method.

strategy, because it may produce a nearly optimal solution, is to use a sequence of functions F_p , $p \in \{0, 1, \dots, r\}$, which lie between F and F^* . They are defined to have the properties

$$F_0 \equiv F^*, \quad F_1 \equiv F$$

and

(5)

$$\forall \bar{\theta} \quad p \geq q \Rightarrow 0 \leq F_p(\bar{\theta}) - F_q(\bar{\theta}) \leq (p - q)\epsilon.$$

Figure 4 shows an example of how to alter the h_i functions to achieve this.

Corresponding to this family of functions F_p is a family of local optima $\bar{\theta}_p$, $0 \leq p \leq 1$, obtained as p increases from 0 to 1. This is obtained from continual hill-climbing on a changing function F_p (rather like keeping a rubber dinghy on the crest of a moving wave). It can be shown that this strategy broadens the search for a solution to allow increasingly suboptimal solutions until one is found. Once a $\bar{\theta}_p$ is found for which condition (4) holds, then the optimisation process can be halted and the solution is $\bar{\theta}_p$, where

$$p = \min_q \{q \mid F_q(\bar{\theta}_q) = F(\bar{\theta}_q)\}. \quad (6)$$

If p is only slightly greater than 0 (i.e. F_p is only slightly non-convex) then it can be shown that the cost $F(\bar{\theta}_p)$ is only slightly greater than the cost of the true global minimum of F . In fact we can find an upper bound ϵ_p on this extra cost, such that

$$p \geq q \Rightarrow \epsilon_p \geq \epsilon_q$$

and

(7)

$$\epsilon_p \rightarrow 0 \quad \text{as } p \rightarrow 0 \quad (p \in [0, 1]).$$

See the appendix for a proof. It is in this sense, then, that the method of 'graduated non-convexity' described above can find a good sub-optimal solution.

4. Results

The methods described in this paper – construction of a convex envelope and graduated non-convexity – have been used to enforce weak continuity both in an intensity array and over chains of strokes. The resulting algorithms have been programmed on an emulator of a parallel array pro-

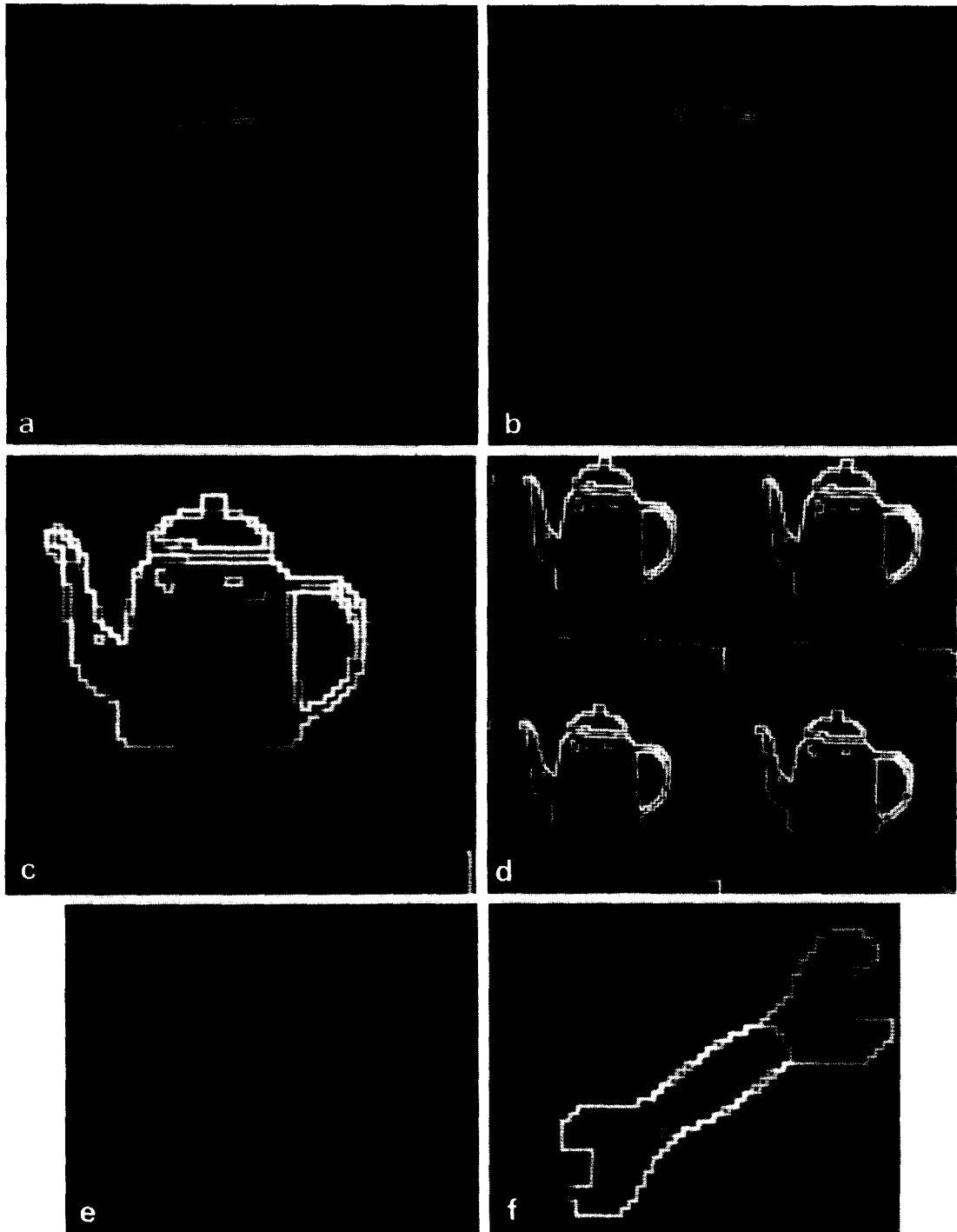


Fig. 5. (a) A teapot, 64×64 resolution, 64 levels of grey. (b) Weak continuity is applied to the picture in Figure 5a, using the method of convex envelopes and graduated non-convexity. (c) Differentiating Figure 5b gives the edge strengths (positions and sizes of the discontinuities). (d) A sequence of edge-strength maps (left to right, top to bottom) in order of increasing penalty for breaking the continuity constraint. In successive frames, the penalty c^2 is doubled. The last frame is the same as Figure 5c. As the penalty increases, the edge map becomes less detailed – fewer discontinuities survive. (e) A spanner. (f) The spanner's edge map (as Figure 5c).

cessor (Blake 1982) and the results for the intensity array are shown in Figure 5. In each case the parallel execution time on CLIP4¹ is about 1.7 seconds.

5. Conclusion

The least-disturbance principle has been proposed for formulating constrained labelling as an optimisation problem. It is applicable when an initial estimate of a labelling is to be adjusted to comply with given constraints.

Some labelling problems, for instance assembling strokes into straight lines and labelling intensity discontinuities, involve weak constraints – constraints which hold almost everywhere. The resulting optimisation problem is non-convex. The non-convexity cannot be removed simply by a more apt expression of the problem; it is a direct consequence of the semantics of the task. Transforming the problem, by construction of convex envelopes, makes the solution accessible, under certain circumstances. Where this fails, the graduated non-convexity method is a good strategy for reaching a suboptimal solution. The method has been used with an emulator of a parallel array processor to impose weak continuity on an intensity array, for segmentation, and to arrange chains of strokes into straight lines.

Appendix. Suboptimality in the graduated non-convexity method

We are given a family of functions F_p , $p = 1, \dots, r$, such that

$$F_0 \equiv F^* \quad \text{and} \quad F_r \equiv F \quad (\text{A1})$$

and also $\exists \varepsilon$ such that:

$$\begin{aligned} \forall \bar{\theta} \quad p \geq q \Rightarrow \\ 0 \leq F_p(\bar{\theta}) - F_q(\bar{\theta}) \leq (p - q)\varepsilon. \end{aligned} \quad (\text{A2})$$

¹ 130 iterations, based on the figures for the CLIP4 design (Duff 1978), not including any extra time required by a serial host machine and assuming that 128 Boolean planes (not 32 as in the design) are available.

As p increases from 0 to r , we obtain a sequence of $\bar{\theta}_p$, the local minima of the corresponding F_p .

We now show that $p\varepsilon$ is a bound on the amount (cost) by which a solution $\bar{\theta}_p$, produced by the graduated non-convexity method, is suboptimal, relative to the desired global optimum $\bar{\theta}'$ of F . It is a property of the graduated method that this solution satisfies

$$F(\bar{\theta}_p) = F_p(\bar{\theta}_p). \quad (\text{A3})$$

The method comprises a sequence of hill-climbing minimisations over the functions F_p , each starting from $\bar{\theta}_{p-1}$ and ending at the local minimum $\bar{\theta}_p$. We have, from (A2),

$$F_p(\bar{\theta}_{p-1}) \leq \varepsilon + F_{p-1}(\bar{\theta}_{p-1}) \quad (\text{A4})$$

and also, as a property of the hill-climbing process,

$$F_p(\bar{\theta}_p) \leq F_p(\bar{\theta}_{p-1}). \quad (\text{A5})$$

From (A4) and (A5), and by a trivial induction, we obtain

$$\forall p \quad F_p(\bar{\theta}_p) \leq p\varepsilon + F_0(\bar{\theta}_0). \quad (\text{A6})$$

But, from (A1) and (A2) we know that

$$F_0(\bar{\theta}') \leq F_r(\bar{\theta}') = F(\bar{\theta}'), \quad (\text{A7})$$

and since $\bar{\theta}_0$ is the *global* minimum of F_0

$$F_0(\bar{\theta}_0) \leq F_0(\bar{\theta}') \leq F(\bar{\theta}'). \quad (\text{A8})$$

Finally, from (A1), (A6) and (A8) we have

$$F_p(\bar{\theta}_p) \leq p\varepsilon + F(\bar{\theta}'). \quad (\text{A9})$$

The extra cost of the suboptimal solution is bounded by $p\varepsilon$. The sooner a solution is found (p small), the closer it is to being optimal.

References

- Blake, A. (1982). PARAPIC language definition, version 2.1. Unpublished report, MIRU, Edinburgh University, Edinburgh.
- Blake, A. (1983). Relaxation labelling – the principle of ‘least disturbance’. *Patt. Recogn. Lett.* 1, 385–391 (this issue).
- Berthod, M. (1982). Definition of a consistent labelling as a global extremum. *Pattern Recognition Conference 1982*, 399–401.
- Duff, M.J.B. (1978). Review of the CLIP. *National Computer Conference 1978*, 1011–1060.

- Hanson, A.R. and E.M. Riseman (1978). Segmentation of natural scenes. In: A.R. Hanson and E.M. Riseman, eds. *Computer Vision Systems*. Academic Press, New York.
- Haralick, R.M., J.L. Mohammed and S.W. Zucker (1980). Compatibilities and the fixed points of arithmetic relaxation processes. *Computer Graphics and Image Processing*. 13, 242–256.
- Hinton, G.H. (1978). *Relaxation and its role in vision*. Ph.D. thesis, University of Edinburgh, Edinburgh.
- Ikeuchi, K. (1980). Numerical shape from shading and occluding contours in a single view. AI Memo 566. AI Lab, MIT, Cambridge, MA.
- Marr, D. (1978). Representing visual information. In: A.R. Hanson and E.M. Riseman, eds. *Computer Vision Systems*. Academic Press, New York.
- Perkins, W.A. (1978). A model-based system for industrial parts. *IEEE Trans. Computers* 27, 126–143.
- Rosenfeld, A., R.A. Hummel and S.W. Zucker (1976). Scene labelling by relaxation operations. *IEEE Trans. Systems Man Cybernet.* 6, 420–433.
- Ullman, S. (1979). Relaxed and constrained optimisation by local processes. *Computer Graphics and Image Processing* 10, 115–125.
- Waltz, D.L. (1972). *Generating semantic descriptions from drawings of scenes with shadows*. Ph.D. thesis, MIT, Cambridge, MA.
- Zucker, S.W., R.A. Hummel and A. Rosenfeld (1977). An application of relaxation labelling to line and curve enhancement. *IEEE Trans. Computers* 26, 394–403.
- Zucker, S., Y. Leclerc and J. Mohammed (1981). Continuous relaxation and local maxima selection: conditions for equivalence. *IEEE Trans. Patt. Anal. Mach. Intell.* 3, 117–127.