1976, p. 148-157
Proc. A.I.S.B Summer Conf.
Univ. of Edinburgh

G.Hinton
Cognitive Studies Program
University of Sussex, Brighton

USING RELAXATION TO FIND A PUPPET

ABSTRACT

The problem of finding a puppet in a configuration of overlapping, transparent rectangles is used to show how a relaxation algorithm can extract the globally best figure from a network of conflicting local interpretations.

INTRODUCTION

The program takes as input the co-ordinates of the corners of some overlapping, transparent rectangles (See figure 1). The problem is to find the best possible instantiation of a model of a puppet. The difficulty is that if we only consider a rectangle and its overlapping neighbours, then each rectangle could be several different puppet parts or none at all, so local ambiguities have to be resolved by finding the best global interpretation. The aim of this paper is to show how a relaxation method can be used instead of the obvious search through the space of all combinations of locally possible interpretations. The relaxation method has several advantages:

1. Using parallel computation the best global interpretation can be found quickly. The time taken is not exponential in the number of local possibilities because combinations are not dealt with explicitly.

2. The computing space required increases only linearly with the number of possibilities, which makes this method better than an exhaustive, breadth-first parallel search, for which there is a combinatorial explosion in space.

3. It produces the best global interpretation, not just a good one as in heuristic search.

All these reasons make relaxation look good as a model of how the brain resolves conflicting low-level visual hypotheses. A conventional, serial A.I. search would be very slow, given the brain's sluggish hardware (Sutherland 1974).

THE PUPPET MODEL

The puppet, which is always depicted in side view, consists of fifteen rectangular parts having the following properties and relationships:

1. Each part has a proximal end and a distal end. The proximal end is the one anatomically nearest to the top of the head. The length of a part, measured along the proximal-distal axis must be greater than its width.

2. The trunk must be wider than any of the upper limb-parts, and each of these, in turn, must be wider than its connected lower limb-part. Also, the head and trunk must be wider than the neck.

3. The head must be greater in area than the neck and the lower limb-parts must be larger than their associated hands or feet.

4. Anatomically connected parts must overlap in the right way. This is defined by specifying zones in each part and then specifying pairs of zones, one in each part, which must or must not overlap. The definition of a satisfactory joint between calf and thigh is shown in figure 2, together with some examples and near misses.

This puppet model is fairly arbitrary, but something more than simple connectivity must be used to exclude cases like figure 3. One way in which people are more flexible (as perceivers!) is that they will allow some relations or proportions to be stretched provided the rest are reasonable. The implications of this will be discussed later.

INCOMPLETE PUPPETS

The data-structure which represents the interpretation of a rectangle as a puppet part is called a percept and has slots which are filled by relations to other percepts. The relations are also represented explicitly by data-structures, which have two slots, one for each of the related percepts. When there is nothing better in the picture, people happily find incomplete puppets, i.e. ones in which some percepts have vacant slots. The program can do the same if it is given some way of evaluating incomplete puppets so that it can avoid poor global interpretations when there are better alternatives. Currently, the best puppet is defined as the one containing the most anatomical relations whilst satisfying the following constraints:

1. No rectangle can be seen as more than one part.

2. No slot can be filled by more than one relation, except for the thigh and upper-arm slots in the trunk, which can have two.

3. No type of part can be instantiated more times than it occurs in the model: e.g. there must not be more than two

thighs.

4. A relation cannot exist unless its percepts do.

This definition has its problems (see discussion section), but for many pictures it is adequate.

For pictures containing perfect puppets plus some optional extra rectangles, the task can be done fairly simply using a technique like the Waltz filter (Waltz 1972). Each rectangle is given a list of locally possible percepts and then, for example, all those potential necks which are not connected to any potential head are deleted. Iterating this process often produces a unique global interpretation and always greatly reduces the search space. Unfortunately, if a part of the puppet is missing, then the correct interpretations of the adjacent parts get deleted until eventually all percepts are wiped out. If the filtering process is weakened sufficiently to prevent incorrect deletion, it loses most of its power, and without an effective filter, the search becomes large, since the possibility that the puppet is imperfect prevents early backtracking.

GROWING A NETWORK OF LOCAL INTERPRETATIONS

Since the potential incompleteness of the puppet makes it hard to rule out any percepts on local grounds, the program uses the alternative approach of ruling them in, starting from local configurations which strongly suggest particular percepts. From these nuclei, it grows a network by attempting to fill the vacant slots with relations to already existing percepts. If this fails, and there are suitable overlapping rectangles, relations to freshly created percepts are used, and the other slots of these new percepts then act as further growing points.

Provided the best instantiation of the model contains at least one nucleus, the resulting network will contain all the required percepts. It will also contain many others and some slots will be filled by several competing relations (See figure 1b). Generally, however, a network grown in this way will be considerably smaller than one consisting of all the local possibilities.

INTERACTIONS BETWEEN PERCEPTS

Parallel processing must create a surfeit of local possibilities to be sure of generating the correct ones, so the time advantages of parallelity are lost unless there is a fast way of eliminating the rest. Simple local competition will not work because a correct percept sometimes has a locally better alternative (See figure 4), but if percepts are also allowed to help one another via their relations, then support may propagate through the network to aid a globally consistent but locally inferior percept (See figure 4 again). A system of this type, in which global patterns emerge from local interactions, is attractive as a basis for Gestalt phenomena, but only if the system quickly reaches a stable state and there is some guarantee that the best pattern emerges.

A certain amount of mathematics is required to show how the interactions can be organised so as to satisfy these conditions. In this brief presentation, I have decided to omit proofs and the precise formalism they require, and concentrate instead on making the flavour of the ideas more easily available to a general audience. A more formal treatment of a similar, independently developed system can be found in Rosenfeld et al. (1975), though their failure to distinguish adequately between preferences and constraints (see below) makes them abandon a linear model prematurely.

PREFERENCE-CONSTRAINT NETWORKS

Finding the best puppet is equivalent to extracting from a network whose nodes are the percepts and relations the best subnet satisfying certain constraints. If the value of a subnet can be expressed as the sum of the preferences for its individual nodes, and if the constraints are equivalent to hyperplanes in the space of possible states (see below), then a relaxation method can be applied. Each node is given an associated real number between 1 and 0 called its credibility. This quantity, which should not be confused with the preference, can be interpreted as the current probability that the node is correct, i.e. part of the best consistent subnet. The constraints are expressed as linear equalities and inequalities between credibilities. For example, n or m is expressed as:

$c(n) + c(m) \geqslant 1$ where $c(n)$ is the credibility of node n.

The credibilities of the nodes can be represented as the axes of a multi-dimensional space. A credibility distribution is then a point in the space, and a constraint corresponds to a hyper-plane. To satisfy an equality or inequality constraint, a point must lie on the relevant hyper-plane or on the appropriate side of it. The states which satisfy all the constraints are called legal states and the region of the space corresponding to them is a convex polyhedron, because it is the intersection of some hyper-planes (equality constraints) and some half-spaces (inequality constraints).

The value of a credibility distribution is the scalar product of the preference vector with the credibility vector. In spatial terms this means that the preferences for the individual nodes define a direction in the space of credibility distributions, and the best legal state is the one furthest in this direction. In general, this will be a vertex of the legal region and can be found using the Simplex algorithm, a standard technique in linear programming (Pierre 1969), or by relaxation which may be better given parallel hardware.

Starting with any assignment of credibilities, each node in turn has its credibility altered so as to minimise the total amount by which the constraints involving the node are violated. It may help to think of the constraints as exerting forces proportional to the size of the violation. These forces cause the credibility to move to an equilibrium position in which they are balanced. Iterating the process for all the nodes, continually reduces the sum of all the violations until there are none left.

This process moves the credibility distribution into the legal region. In order to find the best point in this region, much weaker forces, proportional to the preferences are applied during relaxation. This allows the preferences to pushthe credibility distribution in the best direction, but prevents them from causing significant violations.

We have seen how the best legal state can be found. If it only contains credibilities of one or zero, then the nodes with a credibility of one definitely correspond to the best consistent subnet. If, however, intermediate credibilities occur in the best state, as they do in some as yet ill-defined circumstances, a search must be performed by fixing some nodes at one or zero, and using relaxation to find the values of the others.

APPLYING RELAXATION TO PUPPET PICTURES

A definition of the best instantiation of the puppet was given in the section on incomplete puppets. The constraints listed there can be expressed in terms of credibilities as follows:

1. For percepts corresponding to one rectangle,

   $$\sum c(p) \leqslant 1$$

2. For relations competing for a slot in a percept:

   $$\sum c(r) \leqslant 1 \quad \text{or} \quad \sum c(r) \leqslant 2$$

3. For percepts of a type of part that occurs n times in the model:

   $$\sum c(p) \leqslant n$$

4. For a relation, r, between two percepts, p, q:

   $$c(r) \leqslant c(p) \quad \text{and} \quad c(r) \leqslant c(q)$$

The preferences are zero for all percepts and positive and equal for all relations. Figures 1 and 4 show two of the pictures on which the program has been tried. So far, it has always found the best puppet, but further analysis and testing are required. Notice how sensibly the relaxation method resolves the local ambiguities concerning the trunk in figure 1.

DISCUSSION

The task which has been used to reveal the principles of a relaxation approach, has been simplified in many ways. One easily modified feature is the lack of attention to the angles of the knee and elbow joints. A better puppet model would require that the elbows bent one way and the knees the other, and it should be possible to mobilise such knowledge to make the best puppet emerge. The theoretical interest of this type of constraint is that it is non-local, like number agreement which is problematic for context free grammars (Lyons 1968). The practical solution is to introduce global nodes representing the side of the puppet in view. These side nodes are related to each other by an exclusive-or constraint and each relevant relation is related to the compatible global side node by a material implication constraint. The best instantiation will now have compatible knees and elbows. In some cases this is too severe a restriction, since a broken elbow is better than none. So alternative weaker relations without the extra constraints are introduced. These conflict with the stronger relations, so if they have lower preferences, good elbows will beat poor ones, but poor ones will beat none.

A more serious simplification is that relations and proportions are either definitely satisfactory or definitely not. Intermediate cases could be included in the network, but there are so many of them in a complex scene that the network would become cumbersome, and a lot of computing would have to be done for very little return. The alternative is to start by creating only the good percepts and relations. As the relaxation process is applied, some percepts will emerge with high credibilities and their vacant slots can then be developed. Integrating the growing and the running of the network in this way, effectively prevents a local cue which conflicts with globally better alternatives, from initiating a search for supporting evidence. Its disadvantage is that there is no longer any guarantee that the final global interpretation is the best one, because some of the nodes of the best possible interpretation may never be added to the network.

Finally, in the simple task described, there are only two levels of structure, the puppet and the rectangles. These do not give rise to situations in which higher level knowledge is required to form the right lower level structures. Given an imperfect line drawing with occlusion, however, puppet knowledge may be required to decide which line segments form a rectangle. The obvious way of doing this is to find the best puppet composed of easily found rectangles, and then to search for poorly depicted parts to complete it. A different and novel method is to create many potential rectangles and to set up constraints between them. Rectangles which conflict over the interpretation of a line segment, for example, cannot both depict parts of the best puppet. Instead of using relaxation immediately to find the best consistent set of rectangles, all the possibilities can be used to find potential percepts and relations. The percepts can then be linked by material implication constraints to their corresponding rectangles, thus creating a larger network containing conflicting structures at several levels. By running the whole lot at once, the best puppet can be found, and higher level knowledge can be made to influence the choice of rectangles.

REFERENCES

Lyons,J. (1968) Introduction to Theoretical Linguistics. Cambridge University Press.

Pierre,D.A. (1969) Optimisation Theory with Applications. John Wiley and Sons.

Rosenfeld,A., R.A. Hummel & S.W.Zucker (1975) Scene Labelling by Relaxation Operations. Technical Report TR-379, Computer Science Centre, University of Maryland, College Park.

Sutherland,N.S. (1974) Intelligent Picture Processing Manuscript to be published in: Sutherland,N.S. (ed.) Tutorial Essays in Psychology. Larry Erlbaum Associates. Washington.

Waltz,D. (1975) Understanding Line Drawings of Scenes with Shadows, in The Psychology of Computer Vision, Winston,P.H. (ed.) McGraw Hill: New York.

ACKNOWLEDGEMENTS

I would like to thank all the members of the A.I. community at Sussex, particularly Christopher Longuet-Higgins and Aaron Sloman, for their helpful criticism and useful suggestions.

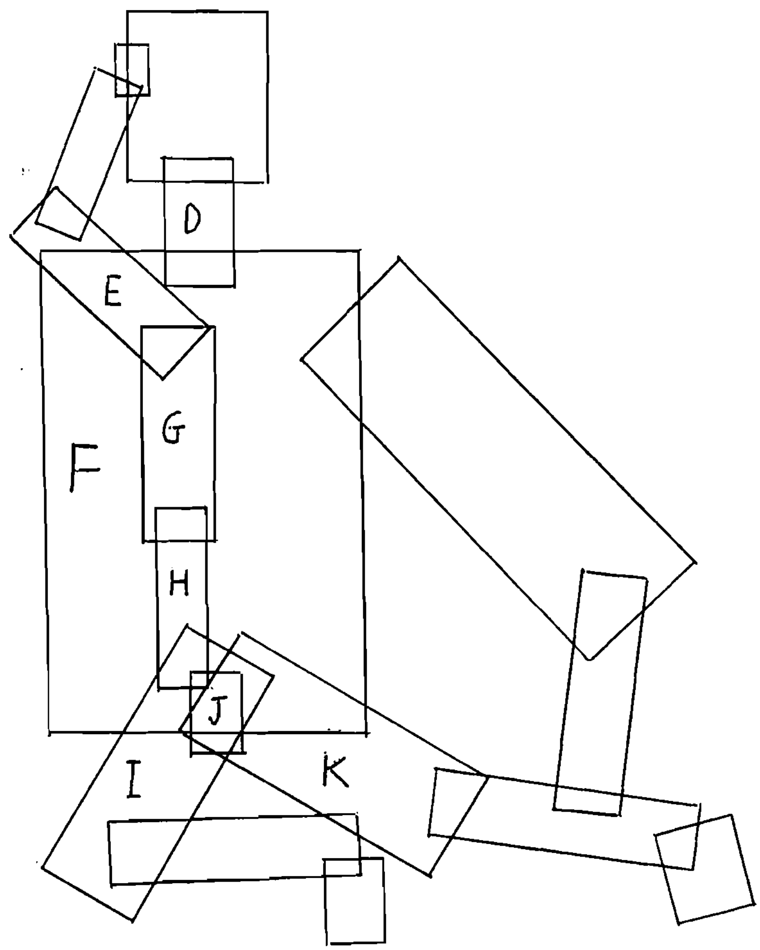


Figure 1a. A puppet with some extra rectangles.

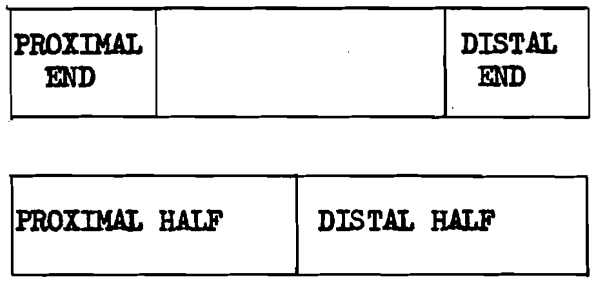
LOCAL HYPOTHESES FOR F BEFORE RELAXATION:

F DOWN TRUNK: NECK I UPPERARM G H I J K THIGH G D E
F UP TRUNK: NECK D UPPERARM G D E THIGH G H I J K
F DOWN HEAD: NECK D

AFTER RELAXATION:

F UP TRUNK: NECK D UPPERARM G E THIGH K I

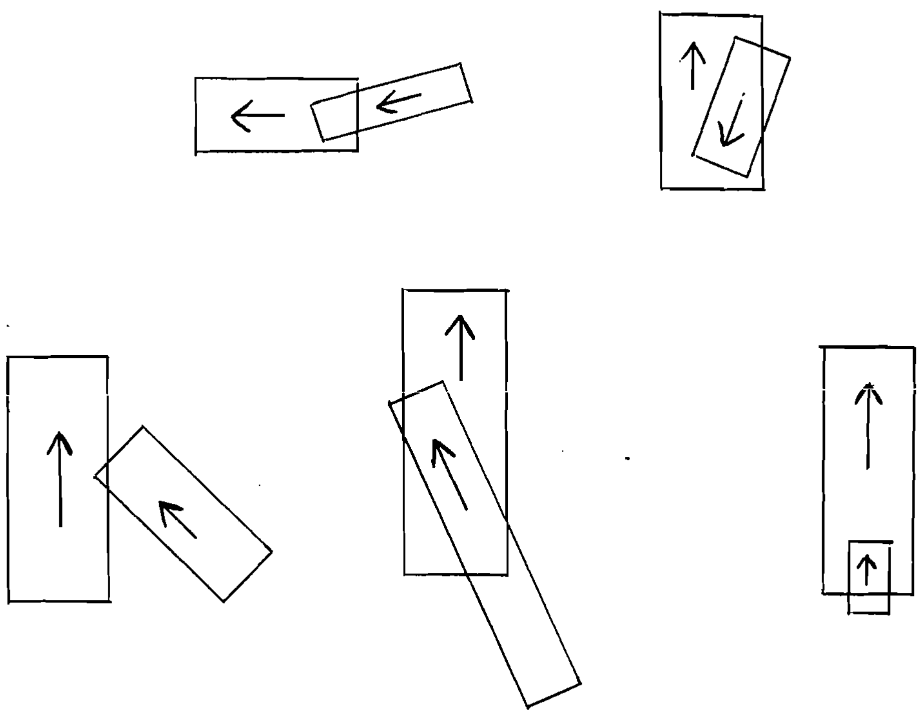
Figure 1b. The local hypotheses for a rectangle before and after relaxation. The second column indicates the direction in the picture, in which the proximal end of the percept is facing. Directly related percepts are shown after the colon.



<u>Figure 2a. Four zones of a percept</u>.

CALF	THIGH	OVERLAP?
PROX. END	DIST. END	MUST OVERLAP
PROX. END	PROX. HALF	MUST NOT OVERLAP
DIST. HALF	DIST. END	MUST NOT OVERLAP

<u>Figure 2b</u>. A table showing the definition of the knee relationship in terms of zone relationships.



<u>Figure 2c</u>. Two examples of a satisfactory knee joint (top) and three near misses. An arrow indicates the distal → proximal direction. The thigh is always the wider of the two.

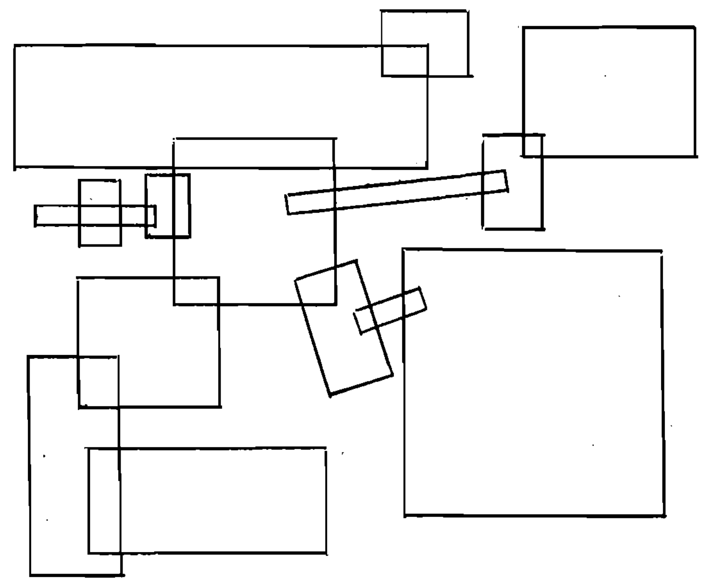


Figure 3. A configuration of rectangles with the same connectivity graph as a puppet, but with different relations and proportions.

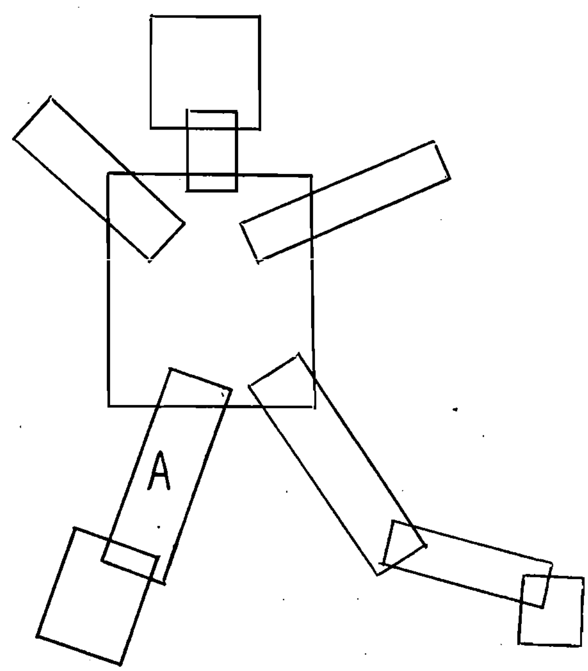


Figure 4. Relaxation picks out the interpretation of A as a thigh even though a calf is a locally better alternative.