RECOGNITION USING SEMANTIC CONTRAINTS

M.O. Shneier
Department of Artificial Intelligence
University of Edinburgh
Edinburgh, Scotland

Abstract

This paper describes a means by which a scene can be split into named subparts, and these subparts can interact to produce an interpretation which is a description of the objects present in the scene. The procedure is in two parts - a set of labels is assigned to each element of the scene, and these sets are decreased in size by examining the allowable relationships between named elements. An application of the procedure to the domain of surfaces produced by the ranger at Edinburgh University is presented.

1. INTRODUCTION

This paper provides an algorithm with general applications in segmentation and classification. In the recognition domain it uses explicit models of bodies occurring in laboratory-created scenes, the models being specified by surfaces and the relations between them. In its operation it displays some of the features which have been found to be useful in analyzing naturally-occurring scenes (Yakimovsky and Feldman (1973)). The justification for presenting this approach is that it makes better use of the information available in the scene than earlier systems, and performs less search. The purpose of the system is to interpret a "scene" by attaching to each input scene element at least one (and preferably only one) name - the name of the body whose image the scene element is a part of. An example of the use of the algorithm is presented in section 2.3.

Each body model is constructed from surface primitives, which may appear in many different models. Relations holding between the primitives in each model serve to differentiate bodies, and play a large part in the recognition, which involves three tasks. (1) Scene elements are paired with model primitives, using a matching function. (2) A set of body names is assigned to each pair, using a mapping from primitives to name sets, and establishing an initial confidence for each assignment. (3) Relations holding between surfaces in the scene are used to filter the sets of names to try to find a unique name for each pair, or, if this is not possible, the set of names with highest global confidence.

The algorithm presented here has a fairly broadly-based history. The use of relational structure was demonstrated by Barrow and Popplestone (1971), and provided with a formal basis by Barrow, Ambler, and Burstall (1972). Consistency in such networks of relations is discussed by Mackworth (1977). The filtering method of Waltz (1972) is a precursor, and a special case of the present algorithm. Waltz provided a "sufficient" enumeration of legal relationships for a subset of line drawings under local restrictions related to their junctions, and a procedure for interpreting such drawings. He dealt only with deterministic confidences - an interpretation was either possible *or* impossible, and all possible interpretations were equally likely.

The use of confidences in assigning interpretations to scenes has become more prevalent recently. Yakimovsky and Feldman (1973) use Bayesian statistics to analyze real scenes, while Hinton (1976) makes use of confidences in his relaxation method of finding the best instance of a puppet in a scene containing a number of overlapping transparent rectangles. Relaxation labelling methods with various confidence relationships between labels and items have been treated formally by Rosenfeld, Hummel, and Zucker (1975).

Barrow and Tenenbaum (1976a, 1976b) have independently developed a system similar to this one, but designed for analyzing natural scenes. Items in the scenes are assigned a set of names, each with an a priori confidence. Relations are then applied between items in the scene to constrain the name sets, and to promote the confidence of the most likely global interpretation.

2. A REAL SYSTEM

A system is being implemented to recognize known bodies appearing in scenes. Solutions to the problems of occlusion, shadowing, and multiple occurrences of a single body type which work within the greater framework of the algorithm are presented.

Models of the bodies are created in a learning phase, and serve as a database for recognition. The input is surface information obtained by means of the Edinburgh Ranging System (Popplestone et al (1975), Popplestone and Ambler (1977)). Because the system uses a triangulation technique, points in the scene which are not in the "line of sight" of either of the bases of the triangle are not visible.

The system uses primitive items which are fragments of planes or cylinders (e.g. a 5 inch square), and the relations between them to describe objects.

2.1 MODELS

An object to be modelled is placed on a turntable and scanned. It is then rotated to show previously hidden surfaces and scanned again. The views are merged, and a single 3-D set of surfaces is obtained, which encompasses the object. For each surface, certain parameters are calculated (e.g. curvature and extent), and used to form a description of the surface to be stored in the database. At the same time relationships holding between the surfaces are worked out, and added to the model.

The modeller outputs three things:

1. A set of primitives, or abstracted surfaces. These contain a description of the kind of surface they represent, and the names of the models they are-associated with.
2. A set of relations. These are legal correspondences, specifying allowed relationships between the primitives (e.g. adjacency). A relation consists of a name, which should be that of an executable function, a list of arguments, which are primitive surfaces, and an expected value linked to the models for which it holds.
3. A body model, consisting of a name, which is user-supplied, the primitives which make up the model, and, for each of these, the relations which involve it and which have been worked out.

An item to be added to the database does not necessarily go in as a new entry. If its characteristics match an element already in the database, the resident element is merely updated to reflect any new information. For instance, a primitive may have the same form for many different models, in which case the only new information will be the name of its associated body.

2.2. RECOGNITION

In the following, the "set of support" is a global list of models which have had primitives assigned to them.

The algorithm can be divided into three stages. Step 1 is concerned with scanning the scene and getting it into a form compatible with the internal representation. The second stage assigns primitives to models in steps 2 to 5 using relatively loose criteria. If there is any doubt, a primitive will be added to a candidate model rather than being left out. Finally, step 6 has the responsibility of weeding out less favourable assignments. It is better fitted to do this because it has more complete knowledge of the situation than earlier steps.

1. The scene is scanned to form surface descriptions. These are matched against primitives in the database, a note being made of which surfaces are occluded or shadowed. A single surface may match, and be associated with, several primitives, in which case the primitives are called linked. Occluded and shadowed surfaces match all compatible surfaces larger than themselves.

   The surface information is associated with each matching primitive since it will be needed for working out the relations.

2. Assign those primitive/surface pairs whose primitives appear in only one body model to that model. If two assignments to the same body model are inconsistent (i.e. do not satisfy the relational constraints) the program assumes that there are two distinct bodies with the same model in the scene, and two copies of that model are set up. These initial assignments have zero confidence, and give rise to the original set of support.
3. While there are still unassigned primitive/surface pairs, repeat steps 4 and 5.
4. One of the primitive/surface pairs which have yet to be assigned to a model is chosen. Since this choice can be important to the efficiency (but not the effectiveness) of the algorithm, it must be carefully made. The pair is chosen which has among its candidates the model on the set of support which currently enjoys the highest confidence, should such a pair exist. Otherwise, the pair whose primitive best matched with its surface in step 1 is chosen.
5. Test the relational constraints. The new primitive has a list of candidate models in which it may appear, and, for each of these, a list of relations it should satisfy if it does belong to that model. The new primitive can apply all those relations whose other arguments are already instantiated. The actual value is the result of applying the function to the surfaces associated with the primitives. This is compared with the expected outcome, with which, for a success, it must match to within a fixed tolerance. Each success enhances the confidence that the arguments belong to the models for which the relation is a constraint, whereas failure reduces confidence only in the suitability of the new primitive to these models. A relation may be appropriate to more than one model, so that results propagate without work being repeated.

On success, the primitive is assigned to all models indicated by the relation. If necessary, a body model is activated and all a relation's arguments are assigned to it.

If no relational tests succeed, assignment of a primitive to all models for which it is a component, with zero confidence, serves as a place marker so that a primitive is not lost. As in step 2, copies of some models may be needed.

When all tests are completed, the models to which the new primitive was assigned are added to the set of support if they are not already members.

6. Apply filtering to the sets of models assigned in the previous steps. An initial confidence in each assignment has been worked out, and, on the basis of this, a way is sought to consistently reduce the number of models assigned to each primitive to one.

Naturally, any unique assignment should be retained, as must any whose rivals all have zero confidence.

When all assignments have positive confidence, the situation is not ao clear. For the best global solution, the consequences both of deleting and retaining each assignment should be followed in their contexts to a stable solution, and that, with highest overall confidence accepted.

Instead of this complete tree search, the initial confidence has been allowed to act as a predictor of the final outcome in cases where one assignment has overriding confidence. Where confidences are nearly equal, this is not justified, and all assignments are considered as ambiguous identifications.

Clearly, deleting a primitive from a model affects the confidence of all primitives related to it. Their confidence must be reduced by the amount by which the deleted primitive confirmed their presence in the model. This is the way in which the filtering is accomplished, confidence reductions causing further deletions elsewhere, until no further changes are possible. Without a backtrack mechanism as described above, the best solution is not guaranteed.

Linked primitives are treated as multiple assignments for the purpose of conflict analysis, and are deleted if a better match is found with one of their links.

The process cycles while it is still possible to adjudicate between assignments, there being a smaller number of primitives at each step.

When all possible assignments have been made, the models remaining are assumed to be those of the bodies in the scene.

Problems with occlusion, shadowing, and multiple appearances of the same body type are dealt with by a mixture of confidence ranking and filtering. Once a surface has been found to be occluded or shadowed, it can be matched with a constrained set of primitives larger than itself, with a reduced confidence. The linking mechanism will then handle the identification in the course of the algorithm. Linking also provides a way of incorporating ambiguous descriptions. Multiple instances of models call for a splitting of primitives between copies of the models when relational tests fail.

The example shows some of the processes in action.

2.3 EXAMPLE

Let the bodies shown in Figure 1 constitute the world model and let the relation of relative-angle be defined between the surfaces making up the bodies. Note that while we expect the relation to have been defined between for example surfaces A-C, A-D, and D-C in figure 2, the surfaces B-D, and B-C would only coincidently be so related if there were some body in the database for which such a relation held. Suppose the scene depicted in Figure 2 has been scanned from a viewpoint in front of the scene, and the following list of surface-primitive matches has been made (The underlining indicates linked primitives).

[(G 2) (H 3) (K 5) (L 4) (J 1) (A 7)
(C 6) (C 8) (B 6) (E 6) (D 6) (D 8) (F 7)]

Assigning primitives which point to only one body:

BI gets (G 2) (H 3)
B2 gets (L 4) (K 5)

Since, in the scene, G-H and K-L are, as expected, at 90 degrees to each other, applying tests between the two primitives assigned to the same body checks out consistently (if not two copies would have been made). (L 4) and (K 5) will probably not match exactly, since (K 5) is occluded.

With only B 1 and B2 already having assigned primitives, the only possible primitive to choose next is (J 1), which has BI and B6 as candidates. A match with BI (and not with B6) causes assignment of (J 1) to BI, and not to B6. At this stage the bodies on the set of support, can suggest no further action, so, since not all primitives have been accounted for, a new choice must be made. Suppose (A 7) is chosen, activating B3 and BA.

There being no tests to be made, the primitive is assigned to both bodies with zero confidence, and B3 and BA are added to the set of support.

These bodies want a primitive of type 6 or 8, and suppose they decide on (C 6). B3 and B5 are activated, and a match is found with B3, so B5 is deleted. Note that we now have increased confidence in (A 7) as a member of B3.

If the next primitive examined is (C 8) we again find a match with BA. Now we have increased confidence that (A 7) belongs to B4.

Looking for another primitive, we might find (B 6), which activates B3 and B5, but matches with neither. An assignment is made to both with zero confidence, a second copy of B3 (which T will denote B3') being set up.

(E 6) is consistent with (B 6) belonging to B3', and is assigned to B3'. (D 6) activates both copies of B3, and B5, matches with (C 6) and (A 7) in B3, so is added to B3. It is not added to B5, nor to B3', since it fails to match. (F 7) matches B3' and can be assigned.

Now (D 8) is considered, and matched with BA and B5, giving the final list of candidates before the filtering as Bl B2 B3 B3' BA B5

The primitives assigned to more than one body are (D 8), (B 6) and (A 7). (D 8) has a much better match in BA than in B5, so is deleted from B5. As a result, the confidence that (B 6) belongs in B5 is reduced to zero. (B 6) now has zero confidence in B5, but has found a match in B3. Thus (B 6) is removed from B5, which can be deleted since it has no more primitives assigned to it. (A 7), however, has a significant confidence in both B3 and B4. It remains associated with both these bodies, pending further filtering.

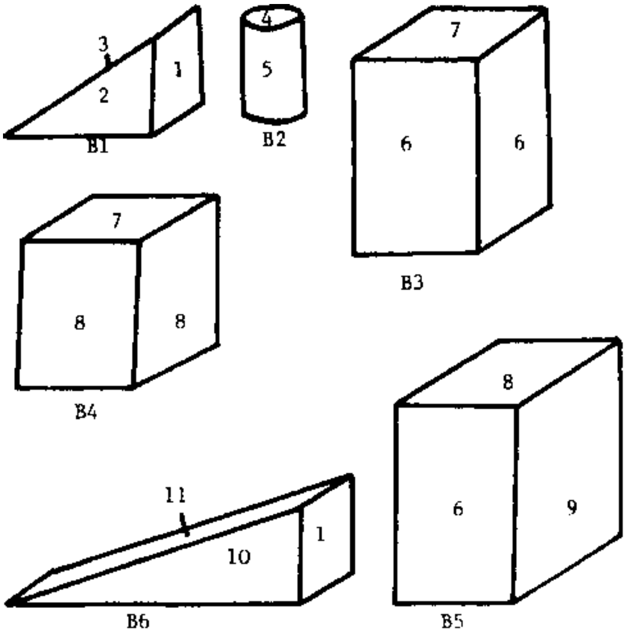
It is discovered that there are linked primitives in B3 and BA. The confidence levels in each are much the same, so nothing can be deleted, and an ambiguous result is announced.

There being nothing else to examine, the final set of bodies recognised is:

BI, B2, B3, and (B3 or BA)

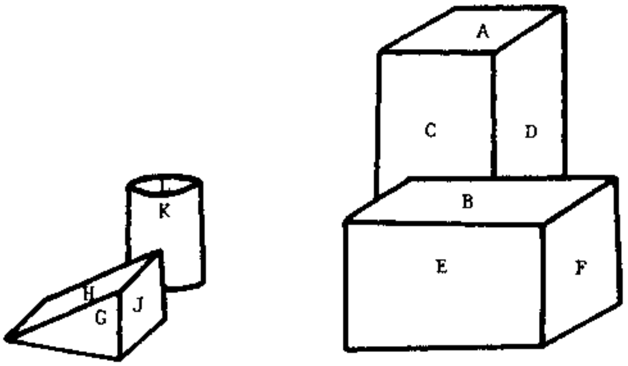
3. DISCUSSION

The system described is in an advanced state of implementation. The modeller is fully operational, while the recognizer works on simple scenes containing single objects.



Primitive	Candidate Models
1	B1, B6
2	B1
3	B1
4	B2
5	B2
6	B3, B5
7	B3, B4
8	B4, B5
9	B5
10	B6
11	B6

The World Model
Figure 1



Example Scene
Figure 2

Symmetry in bodies is handled uniformly, with only one match being found with a model, rather than all isomorphic matches. This is because primitives can occur only once in any model, although instances can be related to several surfaces, and can take up relationships with other instances. For example, a cube model would contain a single square plane primitive, with relations like PERPENDICULAR-TO holding between some of its instances.

Using confidence levels rather than deterministic constraints makes reducing the size of the name sets more difficult. While it is easy to add new names on the basis of even the flimsiest evidence, a deletion must be strongly indicated before the risk of discarding a name can be taken.

The representation makes models easy to learn, and allows great flexibility in the kinds of information known about individual models. Relations which are useful in describing one class of bodies may be unsuitable for others, or information not directly relevant to recognition may be needed for later processing. Since the model indicates which relations to apply, different models can be described in different terms, but instances will still be recognized in a single scene.

Given a vision system with some mobility, it might be possible to disambiguate scenes by moving to another viewing angle, and looking again. Only the particular bodies of interest need be examined, their absolute positions being known. The candidate models can be restricted to those in the ambiguous set, cutting down the amount of filtering needed. This ability to work "top down" from the model to its instances is obtained by restricting the system to look only for primitives needed by the particular model, and only applying relations suggested by the model. It has applications in automatic assembly when particular parts need to be searched for.

Clearly, the method is highly dependent on the amount of information available. The more contraints that can be applied, the fewer ambiguities will result. The efficiency is a function both of the number and type of constraints, and of the way they are represented. It can be conjectured that, given sufficient information, there would be no need for any search.

REFERENCES

H.G. Barrow, A.P. Ambler, and R.M. Burstall (1972) "Some Techniques for Recognizing Structure in Pictures." Frontiers of Pattern Recognition (ed S. Watanabe), Academic Press, New York.

H.G. Barrow and R.J. Popplestone (1971): "Relational Descriptions in Picture Processing." Machine Intelligence 6 (eds B. Meltzer and D. Michie). University Press, Edinburgh.

H.G. Barrow and J.M. Tenenbaum (1976a): "MSYS: A System for Reasoning about Scenes." SRI Technical Note 121.

G. Hinton (1976): "Using Relaxation to Find a Puppet." Proc AISB Summer Conference, Edinburgh.

A.K. Mackworth (1977): "Consistency in Networks of Relations. AI Journal vol 8, No 1.

R.J. Popplestone and A.P. Ambler (1977): "Forming Body Models from Range Data." DAI Research Report, University of Edinburgh, to appear.

R.J. Popplestone, CM. Brown, A.P. Ambler, and G.F. Crawford (1975): "Forming Models of Plane-and-Cylinder Faceted Bodies from Light Stripes." Proc 4th IJCAI, Tblisi.

A. Rosenfeld, R.A. Hummel, and S.W. Zucker (1975) : "Scene Labelling by Relaxation Operations." Technical Report 379, Computer Science Centre, University of Maryland.

J.M. Tenenbaum and H.G. Barrow (1976b): "Experiments in Interpretation Guided Segmentation." SRI Technical Note 123.

D.L. Waltz (1972): "Generating Semantic Descriptions from Drawings of Scenes with Shadows." Ph.D. Thesis, AI Lab Report TR 271, MIT.

Y. Yakimovsky and J.A. Feldman (1973): "A Semantic-Based Decision Theory Region Analyser." Proc 3rd IJCAI, Stanford University.