INTERMEDIATE DESCRIPTIONS IN "POPEYE"

David Owen
Cognitive Studies Programme,
University of Sussex,
Brighton, BN1 9QN England.

ABSTRACT:- Some ideas are presented, derived from work on the POPEYE vision project, concerning the nature and use of different kinds of intermediate picture descriptions. It is suggested that there are "natural elements" in terms of which stored models should be defined and that it is of prime importance to search for those intermediate picture descriptions which are most characteristic of the expression of such elements.

INTRODUCTION

The POPEYE project is concerned with the interpretation in the domain of letters and words of pictures of the kind shown in Fig.1 One of the preoccupations of the project has been the identification of those intermediate descriptions of the picture data which best facilitate the interpretation of the scene from which the data has been derived. An intermediate description corresponds to the identification of a picture object. For example in the POPEYE program contiguous collinear sequences of dots are explicitly represented as "line" data-structures, and pairs of collinear and overlapping lines are explicitly represented as "picture bar" data structures. There are many such objects which may be identified in the picture, corresponding to the representation of "objects" and relations between objects in the different domains involved. (The different domains have been discussed in Sloman et al. 1978).
What follows is a discussion of some emerging ideas concerning the significance of different kinds of picture object and their relation to letter models. An attempt is also made to relate these ideas to the analysis of 3-D polyhedral scenes.

A PARTICULAR VIEW OF LETTERS

A letter is taken to be an abstract object defined by a set of relationships between a number of strokes, themselves abstract objects, and for the current purpose it is only necessary to consider those letters which comprise straight strokes. The important distinction between this kind of description of a letter and a description in some "expressive domain" was made by Clowes (1971). It is also important to distinguish between those characteristics of the representation in the expressive domain which express important properties of the abstract description, and those which are artefacts of the medium of expression.

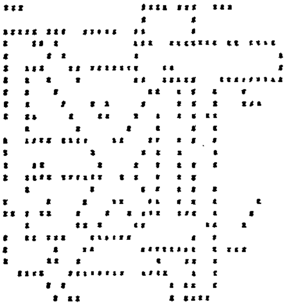


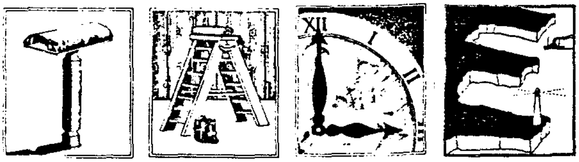
Fig. 1

A 2-D representation of a letter may be regarded as representing two kinds of entity, namely strokes and relations between strokes. Further, it is particular properties of strokes which are of significance and the relationship between two strokes may be described in terms of the values of some simple functions (E.g. difference) defined over the stroke properties. A letter

prescribes for a fixed number of strokes the relative values of the stroke properties by specifying the values of the set of functions defined over them. In such a view of letters, a stroke may be regarded as for example an n-tuple of property values including orientation, length, and position of endpoints.
The significance of using this kind of abstract representation of a letter is that it is independent of the way the values of the n-tuple and any consequent relationship with other n-tuple values are represented. Secondly the functions which describe the relations may be continuous, so that in any representation of a relationship between strokes, the accuracy with which it accords with that prescribed in a letter definition may be measured.

Letter Depictions:-
What is required of the depiction of a stroke is that it should express a particular n-tuple of properties so that a collection of stroke depictions expresses relative values for the properties which may accord with the definition of a letter. Any picture object from which an approximate major axis can be found will fulfil this requirement and some examples are given in Fig.2, of the different ways in which an axis may be defined.

Fig. 2
(From:-
Earnshaw)



The relative values of the properties expressed by a collection of stroke depictions need not conform accurately to those prescribed by a letter definition, for the letter to be recognisable and Fig.2 includes some examples in which the relative values of orientation, length and endpoint positions vary considerably from those of the "ideal" letter they depict. In some of the examples a relationship is not accurately expressed because the corresponding properties are only approximately expressed by the stroke depictions.

A PARTICULAR VIEW OF LETTER RECOGNITION

It may be argued that the underlying theoretical framework of a mechanism which is to interpret a picture in terms of letter depictions has two parts. The first is the recognition of instances of the expression of strokes; the second is a search among those instances for sets of strokes for which the relative values of the properties of the set members conform to within acceptable tolerances to those prescribed by a letter definition.

It is the identification of the two types of task in the underlying theoretical framework which is significant for the choice of intermediate descriptions. They separate the two areas in which the expression of two different types of entity have the potential for great variation, giving rise to the variety of ways in which letters may be recognisably represented (Fig.2.). The first entity is the n-tuple of properties which characterise a stroke, and the second the constraints between sets of strokes corresponding to a particular letter.

One of the implications of adopting such a model is for the relative importance of different kinds of picture object which may be found in the picture. Of prime importance are those which capture instances of the expression of a stroke which in the case of the POPEYE domain is Picture Bars, parallel and overlapping pairs of lines. From them values for all the properties of a stroke may be obtained, and the relative values for different strokes may then be used to address letter models.

The other picture objects which are available in POPEYE, for example line-junctions, are a manifestation of the expression of a precise relationship between between two such strokes. As such they are vulnerable; small changes in the relationship they express will cause them to disappear , without a similarly large effect on the recognisability of the total letter(See also Brady 1978). Their role then, should be as heuristics for limiting the search for which implicitly expressed relationships between strokes are of significance.
In some sense strokes have a "stand alone" meaning, a junction is the precise expression of a compound meaning.

In the POPEYE program this approach has been exploited to some extent. Initial attempts were strongly influenced by the "linguistic analogy", with line junctions taken as the language primitives, and so ideas revolved around grammars over the kind of objects shown in Fig.3. However, the discovery of Picture Bars is now of fundamental importance, and the evidence in the form of line junctions is used more in a segmentation role. The letter models however are still based on a notion of a letter being composed of junctions between strokes, expressed in the form of line junctions. An alternative model system, based on the above ideas, is being developed.

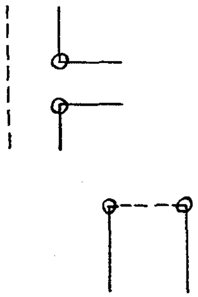


Fig. 3

POLYHEDRAL SCENES

A similar argument applied to 3-D polyhedral scene analysis would suggest that properties characterising a surface would be the counterpart of strokes, and that discovering instances of their expression is of prime importance rather than the manifestations of the expression of a precise relationship between them; (E.g. Fork or Arrow junctions).

Returning to the "linguistic analogy" and considering what Becker(1975) had to say gives this vague notion a little more motivation. Briefly, he argues that speech is generated by a process of "stitching together" appropriate elements from a phrasal lexicon according to grammatical rules. However, the flavour of Becker's paper is an attack on linguists as "frustrated phycisists" for attempting to establish and use grammars only over the primitives of the language, in an attempt to capture the nature of legal sentences in that language. The "principle" which may be extracted from experience with POPEYE, and would appear to have some relevance to 3-D scene analysis, amounts to generalising that criticism of linguistics into the linguistic metaphor in vision, and in particular making a proposal as to the nature of the vision equivalent of the phrases of Becker's lexicon.

Some examples of phrases which Becker gives are as follows:

THIS IS NOT TO SAY THAT

WHAT DOES THIS IMPLY FOR

WE MUST CONCLUDE THAT

Each invokes a meaning in its own right all be it incomplete. Only a few such phrases are required to generate a meaningful sentence, compared with the number of language primitives in the sentence, and this is an important part of the motivation given by Becker for his ideas. The implication is that it is unnecessarily difficult to generate sentences from primitives of the language all the time, that the art which is language acquisition is about learning new phrases and how to "stitch them together" to convey the desired meaning, and that the resulting utterance may be understood in the same way.

The problem with drawing analogies in vision is that it is not obvious what the primitives of the language are (edge features? lines? line-junctions?) and consequently what constitutes a meaningful phrase is equally unclear. In the work of Huffman(1971) and Clowes(1971) in some sense lines are taken as the primitives and line-junctions seem intuitively the most obvious candidates to choose as phrases which include several instances. The intuition arises partly because the affinity between lines is manifest in a most concrete way - they actually touch. Comparing them with Becker's phrases, do they mean anything in their own right?

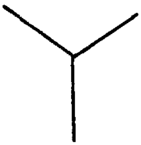


Fig. 4

For example consider the FORK junction in Fig. 4 Clearly in the polyhedral domain this can be taken to "mean" the corner of a cube. However compare this with the following sentence:-

THIS IS NOT TO SAY THAT / ALL MEN HAVE / HAPPY LIVES

which can be taken as comprising three phrases.

Now consider the following three words:-

NOT - MEN - HAPPY

Together they capture more of the meaning of the sentence than any one of the phrases alone because the structure comprising the three words in order captures some mimimal part of the meaning of each of the three phrases. The structure is not of much general use in constructing or analysing sentences since it is a characterisation applicable to only a few sentences. More importantly, unlike each of the three phrases it is ungrammatical.

The suggestion here is that the line-junction of Fig. 4 is more closely identifiable with the three word structure than with a lexical phrase, since together the lines capture some part of the nature of the three surfaces and how they relate, and that the surfaces are better candidates for being the parallels of Becker's phrases.

To continue the comparison, in the same way that only a few phrases are needed to generate a meaningful sentence, only 3 faces of a cube are visible compared with 7 linejunctions. Finally, Becker suggests that speech is generated by "stitching together" appropriate phrases, and it is interesting to note that most people when asked to draw a cube, complete surfaces rather than vertices,i.e. typically a sequence like (a) rather than (b) in Fig.5.

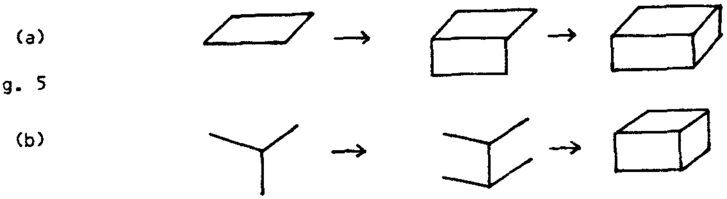


Fig. 5

To return to the comparison with the letter domain, the argument is that surfaces have a "stand alone" meaning and their juxtaposition expresses a compound meaning as an object. Some junctions are particular manifestations of the precise expression of a relationship between two surfaces and as such may or may not capture the compound meaning (E.g. Fig. 6) and are not suitable primtives from which to construct an object model.

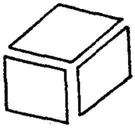


Fig. 6

CONCLUSION

Regarding all vision as involving the addressing of stored models, it is suggested that models should be defined in terms of relationships between "natural elements" which have meaning in their own right, rather than in terms of objects derived from the manifestation in the picture of relationships between such elements. This implies that it is of prime importance to search for those picture objects which are most characteristic of the expression of such "natural elements". It does not imply that other picture objects cannot be exploited, but rather that their usefulness lies in what they imply for the relations between the "natural elements".

ACKNOWLEDGEMENTS

I wish to express my thanks to friends at Sussex for all I have learned from them. Conversations with Steve Draper and Frank O'Gorman have been particularly helpful. Special thanks go to Aaron Sloman for valuable criticism, suggestions and ideas and most of all encouragement.
I am also grateful to Anthony Earnshaw and publishers J. Cape Ltd. for permission to include Fig. 2. This work was supported by the U.K. Science Research Council.

References

Becker J. 1975 "The phrasal lexicon" B.B.N. rep. 3081. A.I. report no. 28 (Cambridge, Mass. Bolt, Berenek, and Newman)

Brady M. 1978 "The development of a computer vision system" Psicologica Recherche.

Clowes M.B. 1971 "On seeing things" Artificial Intelligence,vol.2 no.19.

Earnshaw A. "Seven Secret Alphabets" Jonathan Cape. 1972.

Huffman D.A. 1971 "Impossible Objects as Nonesense Sentences". Machine Intelligence 6. ed.Meltzer B. and Michie D. (Edinburgh University Press).

Sloman A. et al. "Representation and Control in Vision" in Proc. A.I.S.B./G.I Conf. 1978.