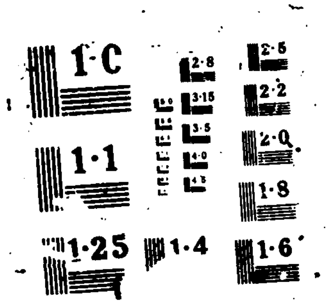
AD-A193 807 SOME ISSUES ON OBJECT REPRESENTATION IN COMPUTER VISION 1/1
(U) ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN
(ENGLAND) P FRETWELL ET AL. JAN 86 RSRE-MEMO-4112

UNCLASSIFIED DRIC-BR-105617 F/G 12/9 NL

END
[illegible]


1·0
2·8
2·5
3·15
2·2
3·5
2·0
4·0
1·1
4·5
1·8
1·25
1·4
1·6

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 4112

TITLE: SOME ISSUES ON OBJECT REPRESENTATION IN COMPUTER VISION

AUTHOR: P. Fretwell, A.C. Sleigh, D.B. Hearn

DATE: January 1986

SUMMARY

It is argued that for most objects in reasonably unconstrained domains previous representations in computer recognition will fail due to the great diversity of appearances of objects within their class. A chair is considered in some detail to illustrate the point. To overcome problems of object variability and other hurdles in recognition a novel type of representation is introduced. This encompasses not only the familiar spatial information but include data on the function and context of what is being recognised. This approach is described and some ideas are given on a "paper" or hypothetical implementation.

Abstract

It is argued that for most objects in reasonably unconstrained domains previous representations in computer recognition will fail due to the great diversity of appearances of objects within their class. A chair is considered in some detail to illustrate the point. To overcome problems of object variability and other hurdles in recognition a novel type of representation is introduced. This encompasses not only the familiar spatial information but include data on the function and context of what is being recognised. This approach is described and some ideas are given on a "paper" or hypothetical implementation.

Contents

1 Introduction

The machine vision activity at RSRE has had emphasis on extracting robust image primitives such as geometric and curved shapes (Sleigh and Hearn 1984), (Sleigh, 1985). This image primitive extraction activity is intended to interface to high level scene understanding process which draw inferences about the likely objects which could generate the image primitives observed. Recently we have been considering the detailed requirements of such a high level scene inference system by supposing that, when introducing a new object to the system, we are wish to make definitions based on human descriptions of objects or scenes, and then apply some form of reasoning (in the loosest sense) to these. There has been considerable work in this area, both among researchers in vision and in the context of general artificial intelligence, (see Chapters 10 and 13 in (Ballard and Brown 1982) for brief survey). Our initial aim is to distil the main elements of this work to establish the necessary requirements a of high level machine vision inference system. We have been chiefly concerned with the fundamental limitations of particular approaches rather than with the detail of implementation. This report is a statement of our thinking so far.

There are several existing methodologies:

- Formation of a relational graph of image primitives, and matching this as a sub-graph of some reference graph defining the world knowledge, see, for instance Faugeras and Price, 1980.
- The closely related approach of using a Frame based approach to represent world knowledge, with some method of filling Slots from image features and applying some form of evidence accrual, (Minsky, 1981).
- Production systems, where objects are expressed by action rules operating on a data base of image features.
- First Order Logic Formalisms where knowledge of objects is expressed as rules operating on facts derived from image features, the basic inference process being that of predicate calculus (modus ponens, conjugation, qualification, etc).
- High order, valued and non-monotonic logics, as for first order logics, but attempting to cope with the incomplete knowledge and default reasoning needed in complex non-closed problems.

Semantic nets and Frame systems attempt to build a taxonomy of architypical objects which can be matched in some way to image features. This usually involves specifying the spatial appearance of the object in terms of generic components, eg generalised cylinders in Acronym, Brooks, 1981 or a richer set of components in more recent Constructive Solid Geometry approaches, Nishihara, 1983. This approach is limited to objects which can be adequately described in spatial terms.

The restriction to spatial description is a serious limitation in complex worlds, since many (perhaps most) objects are best descried in an abstract rather than purely spatial form. For example, is seems reasonable to describe a chair as horizontal surface, supported by four legs connected to a vertical surface. However, many chairs would not conform to this description, eg a swivel chair on wheels, tubular frame chairs with two legs, or an armchair. These differences cannot be adequately captured by allowing a one-legged chair to be 'almost' a four-legged one, for reasons outlined in (Brachman, 1985). This problem can be avoided by specifying the many types of chair as separate concepts, but this is unsatisfactory: cursory examination of chairs indicates that several tens of alternative concepts would be necessary. This will lead to an explosive search problem, since, by the same argument, a similar number of multiple concepts would be needed for each of the subconcepts in a hierarchical description of a an object. If we have such difficulty with a simple object such as a chair, can we hope to be able to make a robust vision system to, say, find its way around an office, distinguish cars from bus shelters, or tanks from cows? We argue that a system based on a purely spatial description of objects is only viable in simple closed worlds.

Rather than attempt to define an architypical chair, we should aim to express the abstract 'essence' of what makes a chair a chair. The number of legs is quite irrelevant, as is the from of connection between the seat and back. In fact a chair must be able to be sat on by one person, supporting a persons back, and this can be achieved in many different ways, not all of them can be represented in a closed world. Instead it appears better to describe a chair as 'something to sit on', going on to define what attributes 'something to sit on' entails, eg stable support, suitable size, strong enough construction etc., only accessing spatial concepts via this sort of abstract descriptive chain. This poses many difficulties, but if they can be overcome the resulting system will have many advantages over purely spatial systems. The definition of a new object will not require explicit description of all possible manifestations, it should have significantly fewer concepts to examine, and will be able to identify unusual instances.

We are not concerned with a system which can automatically formulate its representation of general descriptions from direct experience. Whilst such a system might be conceivable for purely spatial descriptions (although even this is uncertain), it is much more difficult to propose a system which can capture general abstract relationships to perform recognition of generic objects. We suppose the prime concern is to be able to have a dialogue between the system and a human who can 'verbalise' a dictionary describing objects and their subconcepts, and then perform inference on image primitives using the knowledge so described.

We propose and approach based on three elements.

- A human interface using a form of linguistic description of objects which generates a taxonomy of subsumption relationships and attributes based on a Frame concept, except that the taxonomy will be exclusively descriptive rather than assertional (in the sense of (Woods, 1975)) and which is used only to express the prior knowledge about concept relationships.

- A logic system which has access to the taxonomic concepts which enables general relationships to be expressed which would be either impossible or explosive in a taxonomic form (eg. conjoined concepts such as large red bus).
- Search and reasoning which can use the logic system and taxonomy to draw inference about objects from image features extracted from the image by lower level algorithms. This must be able to use image cues to control the logic process, and must be able to cope with incomplete and inconsistent data in a non-closed world.

However, one possible implementation will be discussed which is in sympathy with, but distinct from, the Krypton system (Brachman etal 1983). The presentation will expand on these aspects and discuss an extended example to highlight some of the issues and future problems.

2 Fundamental considerations

Most previous vision systems have used combinations of simple 2D and 3D sub-features to represent objects. These are stored and used as a type of template to match features generated from the image. Although there are lots of variants on this theme the general method can be described as the construction and matching of relational graphs of low level image features

Implementations are generally frame based or based on the idea of relational graphs. So lets see how these two techniques can be used to describe a chair.

As you can see in Figure 1 the relational graph consists of nodes and links. The nodes correspond to the principle features of the chair. The links in-between the nodes describe how the nodes are related.

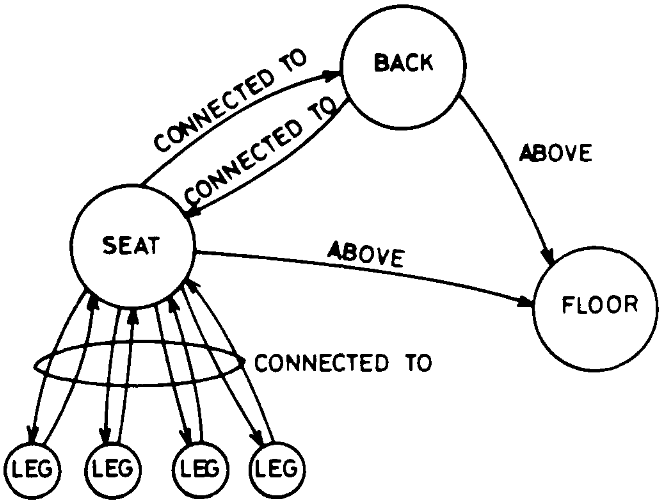
Now what about frames. How do they cope? Consider the Frame representation in Figure 1. The representation include a list of properties that an average chair possesses.

Both these methods need an average or prototypical chair to be effective. But a chair does not have a single all encompassing prototype. Lets see where some of the problems lie in using prototypes to represent knowledge.

Even familiar objects like a chair will have subclasses. There could be office type chairs and the usual four legged variety. Each one of these classes will require a prototype. The advantage of having lots of prototypes is clear - *the more prototypes you've got the more objects you can accurately describe*. However lots of prototypes create problems when your system is searching for a match. Another serious problem is that as more prototypes are added to the system the distinction between them and other object types can become become increasingly blurred. Its also worth considering what happens when the system encounters an object it has not seen before. Also how similar does an object need to be to an existing prototype for the system to identify

WHAT ABOUT REPRESENTING A CHAIR ?

a) RELATIONAL GRAPH



b) FRAME

CHAIR	
IS_ A	: OBJECT TO SIT ON
SUBPART_ BACK :	#1
SUBPART_ SEAT :	#1
SUBPART_ LEG :	#4

Figure 1: Relational Graph and Frame Object Representation

it? And how large does the training set need to be before your system can accurately recognise? So there are problems with the relational graphs and frame approach

What we really want is to have a description that could cover all chairs. So we need necessary conditions for a chair or the "ESSENCE" of the concept "chair".

If we had this we could recognise new and novel forms of chair. Thus the chair with a large helical spring as a support would present no problem to the understanding system. We think that such a description involves defining what function a chair does. Well, it is designed for a person to sit on. A consequence of this is that chair needs a seat and a back that is on a stable support.

We envisage that people will initially input such "function" as well as general spatial and contextual descriptions to the system. This should not present to many problems because although people find it difficult to verbalise shapes, linguistic descriptions of objects in terms of their necessary attributes is well developed. This will aid the difficult knowledge elicitation process.

Using a description of necessary attributes has the advantage that the description is in terms of attributes that are not necessarily dependent on the exact physical form of the object.

Therefore, the Essence representation as it will be called from now on overcomes some of the problems inherent in the low level feature graph approach.

The relationship between function and form has been considered in Lowry, 1982 and by Ingrand for the case of fixture design in Ingrand, 1984. Brady and Agre considered among other things the relationship between the function and form of a set of tools in Brady and Agre, 1984. However of most relevance to the work presented here is the work presented in Adorni et al. 1984, Di Manzo etal, 1985a and Winston et al, 1983. Of special interest is the work in DiManzo etal, 1985 on building function descriptions for image understanding. The emphasis in this paper is on the construction of the functional description and less on the derivation and matching of the functional description to image derived data. This is understandable as this is probably the most difficult part of using function in computer vision. They suggest that functional descriptions can be decomposed into Functional Primitives which can be used to organise complex descriptions according to a conceptual syntax. This is similar to our idea of using a set of common "concepts" that will allow the functional description to be expressed in. Our common concepts correspond quite closely to their functional primitives. The idea of using a set of common concepts to express the information in the Essence descriptions will be described subsequently in this report.

The DiManzo etal paper also describes in outline the approach for mechanism for matching the functional description against an image. This is achieved by expressing the functional description in terms of a set of rules. These rules are manipulated by a rule based expert system. The organisation of the expert system and the way in which it selects parts of the image to be

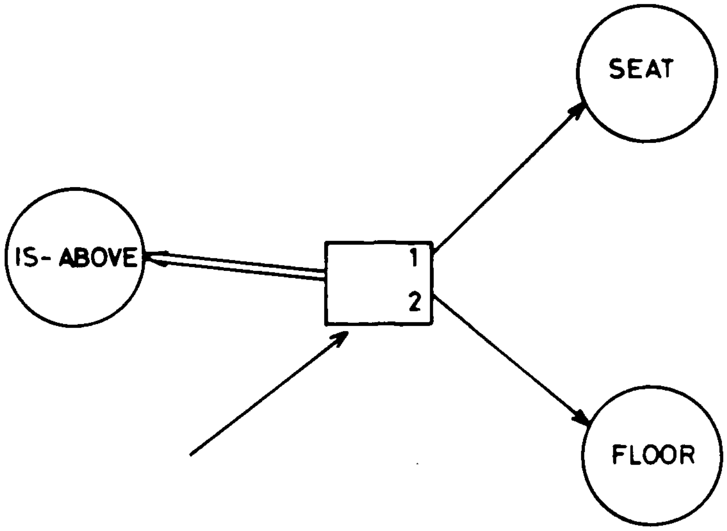


Figure 2: Instance of 'is above' relation

processed are the main thrust of the paper. The algorithms *This is in contrast* to our approach which emphasises the derivation of the functional description and the matching of the functional description to image derived information

A major problem in Artificial Intelligence is that of search. Search in image understanding exists because of the matching of input data to stored data. An advantage of using a generic description like the Essence description is that matching can be guided using information in the image and in the description.

However, the Essence approach is not without its difficulties. The problem of defining and using high level concepts shouldn't be underestimated. Many concepts will be needed to make a working system. This could lead to a large number of inference steps being required for manipulation and reasoning with the concepts.

So our thoughts so far have been that the object representation must be capable of interacting with the user. We feel that the knowledge elicitation process is fundamental to an image understanding system. Secondly, the implementation must be capable of efficiently representing generic objects. In other words the method must be able to efficiently represent many, varied objects. Additionally, the form and organisation of the stored knowledge must aid search and reasoning

3 Essence approach

STYLISED LANGUAGE

CONCEPTS ~ NOUNS

RESTRICTIONS ~ ADJECTIVES

RELATIONSHIPS ~ VERBS

Figure 3: Stylised Language Interface

Lets see how these three points could be developed

Lets take the user interface first. The idea behind this is to allow the user to input descriptions to the system. The proposal is to use dictionary like definitions. These could be represented by verbalised descriptions in a suitably stylised language. Lets take a look at some features such a language would have.

In the language we need to define the essence of the object. This will involve creating and using meaningful sentences. What sort of features would such a language need? We would be describing objects in terms of other objects and concepts like colour for instance. We also need a form of concept restrictions to give us more descriptive power.

For example, it is very inefficient to have a separate concept for every different coloured "bus", much better to have a method of associating the concept bus and the concept "colour". So we need an analogous linguistic entity to an adjective. It can be seen that using adjectives reduces the number of concepts stored in the taxonomy. So we don't need to store a red bus concept and a green bus concept but only need to store a bus concept and a colour concept and have a method of relating these concepts together.

Also we will need to describe the interaction between objects and concepts and that's were the analogy with verbs comes in.

An example of the type of relationship is the ' is above' relationship. The structure of a verb has the form featured in Figure 2. The use of a verb creates a particular structure to represent it and in this particular case the structure relates 'is above' to seat and floor?

So we have seen what sort of features the stylised language would have. This is summarised in Figure 3. How would its be used to form an efficient user interface and how would this sort of description aid search and reasoning.

Its going to be necessary to construct a system that has an efficient way of converting and storing the linguistic descriptions input by the user. Its also going to be necessary to perform search and reasoning with the stored representations. This will require a different and more suitable formalism to that of a knowledge taxonomy. This gives the proposed scheme close analogies with the KRYPTON knowledge representation system, Brachman etal, 1983.

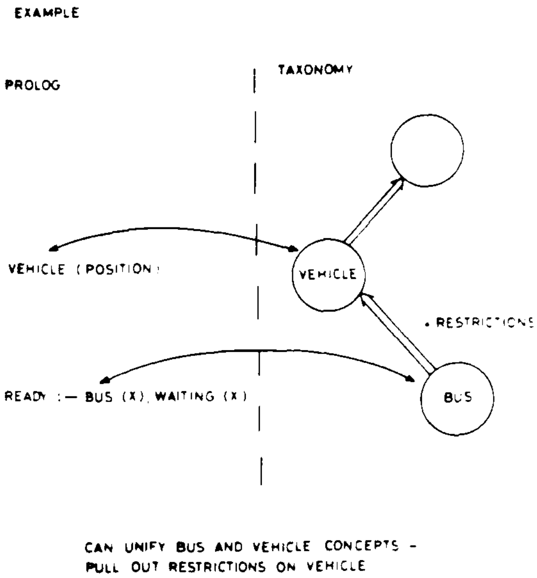

EXAMPLE
PROLOG
TAXONOMY
VEHICLE (POSITION)
VEHICLE
+ RESTRICTIONS
READY :— BUS (X), WAITING (X)
BUS
CAN UNIFY BUS AND VEHICLE CONCEPTS –
PULL OUT RESTRICTIONS ON VEHICLE

Figure 4: Unification of Concepts

An important property of a dual representation structure like this is that it allows the possibility of a very general matching mechanism Two concepts can be matched if they are related by an indirect link such as subsumption An inference system based on this form of matching could access the particular restrictions and specialisations that implicitly differentiate two symbols, and use this in its reasoning process. For example, bus and vehicle would be matched as in Figure 4.

If the symbols defined in the hierarchy where made to correspond to the functors of a PROLOG program for instance, PROLOG unification could take account of the relationship between two concept as a method of unifying two terms even though they have different functors.

Because an inference mechanism can be defined separately from the definition and representation of concepts, there is no need for 'variables' in the concept descriptions. The concept taxonomy is a fixed unchanging set of relationships which is used in the inference process.

Simple reasoning in such a system would be relatively straighforward. However, typical image understanding problems require reasoning with incomplete, incorrect and inconsistent data. You can not assume that your reasoning takes place in a closed world either. Then there is the problem that it will generally be impossible to exhaustively search the problem space.

Because of these three constraints it is apparent that non-standard logic and inference techniques must be used. We plan to address some of these problems in the future However, we will be using first order logic as an initial exploratory tool.

Therefore, the system we are proposing will consist of a static taxonomy of information that is obtained from the user using linguistic type descriptions. Assertional statements about information in the taxonomy are written in a

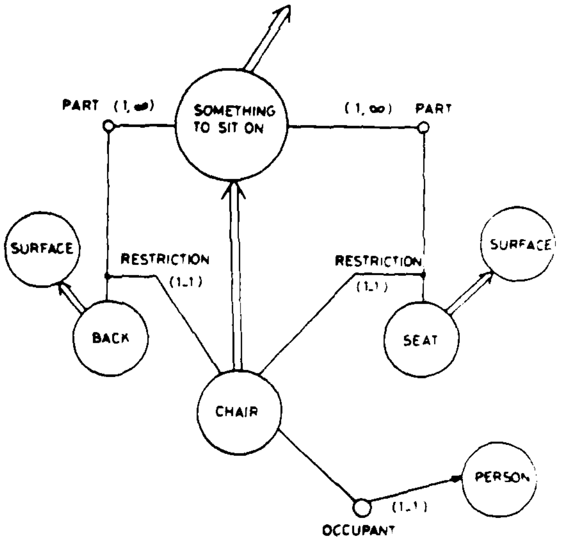


Figure 5: Definition of chair

logic type formalism. This allows the representation greater flexibility in expressive power. An associated element of the system is the facility of a mechanism for efficient search and reasoning with the taxonomic information.

4 Extended example of chair

In an attempt to give a flavour of the approach reconsider the chair example. As stated earlier a chair could be defined as an object for one person to sit on. This leads to the requirement for a seat joined a back on a stable support. Figure 5 illustrates the definition. A chair is a sort of 'something to sit on' concept. Extra conditions are placed on this concept to restrict the number of seats and backs that a chair can have. For instance, the number of backs and seats are constrained to be one only; a chair with no back would be a stool. The 'something to sit on' concept places no restrictions on the number of the number of seats, backs or occupants.

In addition to these restrictions, other attributes of a chair can be defined. For example, relations between parts of the chair and it surroundings are important. These are spatial in nature. The back is connected to the seat, and the chair is supported above the ground. Note that relations of this type are not spatial in the same sense as would be present in, say, a chair model where size and shape would be important factors.

The relationships shown in Figure 6 are shown as instances of the 'joined to' and 'supported above' relations. The generic relations would contain information relevant to their interpretation, such as whether the relation is

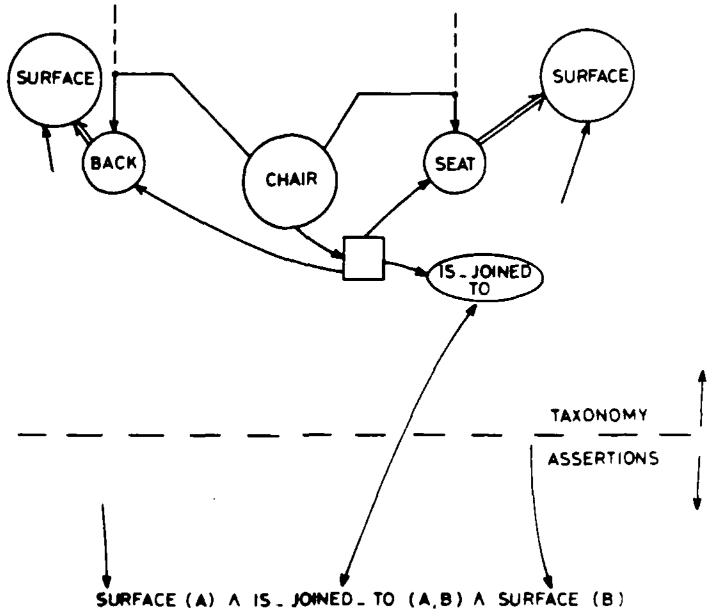


Figure 6: Unification of chair Definition

symmetric, reflexive or transitive The instance structure would just point to the generic relation and its subjects

Consider how the structure in Figure 6 might be used in practice Consider the situation in Figure 6. The ideas embodied in this simple example are capable of being extended to the case where several inference steps are necessary between image features and concepts in the taxonomy

We assume that some low level image processing routines have extracted two surfaces from the image and asserted these into a database contained within the reasoning part of the system In addition the fact that these two surfaces are potentially joined has been deduced from the fact that they share a common side. If the system is looking for chairs in the image, it would look to the database to provide cues. The chair concept requires a back, a seat and the satisfaction of the relations present

In trying to satisfy the requirement for a back and a seat it could unify these with the two surface facts by using subsumption The 'joined' relation on these two surfaces would also be satisfied by the database content

This leaves the 'supported above' relation and the occupancy to be determined. If the system is using the information present in the database as cues, then the result of this cueing would be to attempt to satisfy or to validate these. Image operations or database inference would ensue; any image operations could take place as side effects of the inference

5 Conclusion

We have considered systems that represent objects spatially. It has been demonstrated that this method has serious limitations. System that use spatial as well as more abstract features have been discussed. This has been proposed as a means of overcoming many problems inherent in the purely spatial approach.

The current system we are considering consists of a knowledge taxonomy along with a logic system - this seems to offer many advantages over many previous knowledge representation systems.

Our plans are to develop these ideas into a working system

References

- Adorni, G., Di Manzo, M., Giunchiglia, F., Massone, L., October, 1984, "A conceptual approach to artificial vision", Proc. of the 4th Int. Conf on Robot Vision and Sensory Controls - RoVisSec 4, London.
- Ballard D. H. and Brown C. M., 1982. "Computer Vision", Prentice 1982.
- Brachman, R.J., Fikes, R.E., 1983, "KRYPTON: A Functional Approach to Knowledge Representation", IEEE Computer, 16(10), pp 67-73.
- Brachman R. J., 1985 'I lied about the trees' OR Defaults and Definitions in Knowledge Representation Artificial Intelligence Magazine, Fall 1985, Vol 6, 3, 80-93.
- Brady,M., Agre, P.E., 1984, "The mechanics mate", Proc. of 6th European Conf. on Artificial Intelligence, ECAI-84, Pisa, Italy, 5-7 Sept 1984, pp 79-94.
- Brooks R., 1981. Symbolic Reasoning Among 3D Models and 2D Images Artificial Intelligence Journal, August 1981.
- Di Manzo, M., Adorni, G., Giunchiglia, F., Ricci, F., 1985, "Building function descriptions : Computer vision", Proc. of the 5th Int Conf. on Robot Vision and Sensory Controls - RoVisSec, Amsterdam, Netherlands, 29-31 Oct. 1985, pp 403-412.
- Faugeras, O., Price, K., 1980, "Semantic description of aerial images using stochastic labelling", Proc. ARPA, Image understanding workshop, McLean, VA.:Science Applications, pp 89-94.
- Ingrand, F., Latcombe, J.C., 1984, "Functional reasoning for automatic fixture design", November, 1984, CAM-I 13th. Annual meeting and technical conference, Clearwater Beach,, Florida.
- Lowry, M.R., 1982, "Reasoning between structure and function", Proceedings of the DARPA image understanding workshops, pages 260-264, Science Applications Inc., McLean, VA., September.
- Minsky M., 1981. A Framework for Knowledge Representation in "Mind Design", Ed J. Haugeland, MIT Press 1981 95-128.
- Nishihara H.K., 1983. Recognition of Shapes in Visible Surfaces in "Physical and Biological Processing of Images", Eds O. J. Braddick and A. C. Sleigh, Springer, 1983.
- Sleigh A. C. and Hearn D. B., 1984. An IKBS Approach to Image Understanding, RSRE Memo 3683.
- Sleigh A. C. 1986 "The Extraction of Boundaries Using Local Measures Driven by Rules", Accepted for publication by Pattern Recognition Letters.

- Winston, P.H., Binford, T.O., Katze, B., Lowry, M., January. 1983. "Learning physical descriptions from functional definitions, examples and precedents", M.I.T., Artificial Intelligence Memo.
- Woods W., 1975, What's in a link: foundations for semantic networks. *"Representation and Understanding"*, *Eds Bowbrow, D. G., Collins, A.*, New York: Academic Press.

DOCUMENT CONTROL SHEET

Overall security classification of sheet ....UNCLASSIFIED..

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the box concerned must be marked to indicate the classification eg (R) (C) or (S))

1. DRIC Reference (if known)	2. Originator's Reference Memo 4112	3. Agency Reference	4. Report Security Classification U/C

5. Originator's Code (if known) 778~~5~~400	6. Originator (Corporate Author) Name and Location RSRE St Andrews Road, Malvern, Worcs. WR14 3PS
5a. Sponsoring Agency's Code (if known)	6a. Sponsoring Agency (Contract Authority) Name and Location

7. Title
SOME ISSUES ON OBJECT REPRESENTATION IN COMPUTER VISION.

7a. Title in Foreign Language (in the case of translations)

7b. Presented at (for conference papers) Title, place and date of conference

8. Author 1 Surname, initials FRETWELL, P	9(a) Author 2 SLEIGH, A.C.	9(b) Authors 3,4... HEARN, D.B.	10. Date	pp. ref

11. Contract Number	12. Period	13. Project	14. Other Reference

15. Distribution statement

Descriptors (or keywords)

continue on separate piece of paper

Abstract It is argued that for most objects in reasonably unconstrained domains previous representations in computer recognition will fail due to the great diversity of appearances of objects within their class. A chair is considered in some detail to illustrate the point. To overcome problems of object variability and other hurdles in recognition a novel type of representation is introduced. This encompasses not only the familiar spatial information but include data on the function and context of what is being recognised. This approach is described and some ideas are given on a "paper" or hypothetical implementation.

END

DATE

FILMED

8 88