

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Technical Memorandum 33-693*

*On the Recognition of Complex Structures: Computer  
Software Using Artificial Intelligence Applied  
to Pattern Recognition*

*Yoram Yakimovsky*

PRICES SUBJECT TO CHANGE

(NASA-CR-139363) ON THE RECOGNITION OF  
COMPLEX STRUCTURES: COMPUTER SOFTWARE  
USING ARTIFICIAL INTELLIGENCE APPLIED TO  
PATTERN RECOGNITION (Jet Propulsion Lab.)

N74-29532

CSCL 09B G3/08 44990

Unclas

Reproduced by  
NATIONAL TECHNICAL  
INFORMATION SERVICE  
US Department of Commerce  
Springfield, VA. 22151

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

May 1, 1974

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Technical Memorandum 33-693*

*On the Recognition of Complex Structures: Computer  
Software Using Artificial Intelligence Applied  
to Pattern Recognition*

*Yoram Yakimovsky*

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

May 1, 1974

**Prepared Under Contract No. NAS 7-100  
National Aeronautics and Space Administration**

## PREFACE

The work described in this report was performed by the Space Sciences Division of the Jet Propulsion Laboratory.

Preceding page blank

## CONTENTS

Introduction . . . . .	1
Problem Domain . . . . .	1
Model A: "First Order" Structure . . . . .	2
Model B: Utilizing Procedural Knowledge Through Sequential Classification . . . . .	2
The Sequential Classification Mechanism (Classification Tree). . .	4
Results and Conclusions . . . . .	4
Appendix 1: Assignment Algorithm for Model A . . . . .	4
Appendix 2: Assignment Algorithm for Model B . . . . .	5
References . . . . .	6

## FIGURES

1. Example of steps of analysis of a road scene . . . . .	7
2. Another example of road scene analysis . . . . .	8
3. An example of steps of analysis of an X-ray image of a left ventricular angiogram . . . . .	9

Preceding page blank

## ABSTRACT

An approach to simultaneous interpretation of objects in complex structures so as to maximize a combined utility function is presented. Results of the application of a computer software system to assign meaning to regions in a segmented image based on the principles described in this paper and on a special interactive sequential classification learning system, which is referenced, are demonstrated.

## Introduction

This paper presents principles underlying a software system which interprets relatively complex structures composed of many components that are mutually dependent in their meaning and properties. These principles are, in a sense, a cross-breed between Artificial Intelligence and pattern recognition.

Most of the theoretical works on pattern recognition address themselves to the problem of assigning a label (meaning) to a given object so as to maximize a utility function. The utility function is defined using: (1) the a priori probability of an object to be of a given label, (2) various measured features of the object, and (3) the probability that an object of a given label will have those features.

The research in pattern recognition was mainly aimed at development of: (1) techniques for learning various probabilities of events in the specific problem domain (if not known a priori), (2) ways of making efficient use of features to distinguish between objects of different classes, and (3) various techniques used to assign a meaning to an object so as to maximize the utility functions. However, structures in the real world have a tendency to be composed of many components, each of which is an autonomous object, and the mutual structure is usually very helpful in recognizing them. This is true for speech and written texts (meaningful sentences), as well as for general pictorial information (like various images used in biomedical diagnosis or product quality control).

From all approaches to pattern recognition, the closest approach to ours is the probabilistic grammars (Ref. 1). Both try to parse complex structures, but the mechanism (syntax) which represents knowledge (semantic) is entirely different. The parsing (meaning assignment) algorithm suggested in this paper, even though it may be extended to handle the parsing of certain grammars, was designed mainly to parse structures using real-world knowledge represented using sequential classification mechanisms.

Many researchers in Artificial Intelligence (AI) address themselves to the specific problems arising when attempts are made to design machines which interact intelligently with relatively unconstrained environments. Whenever it was necessary to sense an environment in some way, a real-world pattern

recognition problem arose. However, the techniques applied were usually of a computer science nature, such as: (1) algorithms for searches through trees of states, (2) integer programming and constraint satisfying, (3) logic, (4) procedural knowledge, (5) data bases, (6) special languages, (7) interactive programming, and other such software products. These approaches were sufficiently powerful for some problems, but the absence of probabilistic modeling and the lack of well-defined utility functions resulted in an ad hoc system. There was no effective way to evaluate various decision processes and performance, except by applying the resulting system to various examples and evaluating the output manually. A typical system is described in Ref. 2.

Many of the ideas presented in this paper arose in the context of AI research. The author believes that combining AI techniques with statistical pattern recognition will result in an approach which will be significantly more powerful than each used alone, at least in the area of pictorial pattern recognition.

## Problem Domain

The problem domain to which the specific software system, whose principles are described in the paper, is applied is pictorial pattern recognition. This type of application differs from character recognition in that the objects to be recognized are not line-shaped, but are areas, and are not purely black objects on a white background; they may vary in their light intensity. Examples of analyzing left ventricular angiograms and road scenes are shown at the end of this paper (Figs. 1 - 3). Using various clustering techniques and edge detectors, the raw picture is reduced to a few connected domains (regions) of almost uniform intensity. The reduction is done with conservative enough parameters so that the images of two distinct objects will not be merged into one domain. As a result, the image of a single object is likely to be broken into a few domains. This clustering algorithm is described in Ref. 3. Experiments proved that there is no way to get a perfect segmentation (having one segment per imaged object) using only measurements at grid points without using semantics of the expected scene. Shadows and changes in light reflectance as a result of surface direction change and local textures on the surface of a single object are sometimes more distinguishable than the real boundaries of the object, and as such prohibit a perfect

segmentation without understanding the image. Hence, in relatively early phases with quite a few redundant segments, interpretation of the image is needed.

Suppose that there is a segmented image with  $R(i)$ ,  $i = 1, \dots, N$  domains (regions, segments), and let  $B(i, j)$  stand for the common boundary between domain  $R(i)$  and domain  $R(j)$  if it exists, ( $B(i, j)$  consists of one or more continuous paths). A global interpretation of the scene will mean an assignment of meanings to each domain. Let INT stand for a global interpretation function, then

$$\text{INT: } R(i) \longrightarrow j$$

where  $R(i)$  is an element of the set of all domains, and  $j$  is one of the few possible labels (interpretations) for domains (INT ( $R(i)$ ) can be equal to INT ( $R(j)$ ) when  $i \neq j$ ). The range of labels is defined by the scope of possible objects in the expected image. The reader should remember that INT ( $R(i)$ ) =  $j$  should read, "the domain  $R(i)$  is part of the image of the real-world object of meaning  $j$ ," because often the domain is only part of the image of the object. Our task is to find an interpretation function INT such that the probability

$$P\{\text{INT}(R(i)) \quad i=1, N \mid \text{measurements of segmented image}\} \quad (1)$$

which is our utility function, is maximized over all possible INT(.). Expression (1) can be read: the probability of the event: "R(i) is part of the image of an object of label INT( $R(i)$ ) for  $i = 1, \dots, N$ , given all measurements of the segmented image." The measurements are the gray level readings at the various points of the image and the boundary lines which segment the image.

The process of looking at each individual segment as it stands alone and determining the object for which it is part of the image is the classical pattern-recognition problem. Assigning meaning to each domain independently of interpretation and features of other domains will be equivalent to finding an interpretation function which maximizes the following utility function:

$$\prod_{1 \leq i \leq N} P\{\text{INT}(R(i)) \mid \text{measurements inside } R(i)\} \quad (2)$$

over all possible interpretation functions. However, by considering individual objects one at a time, as if they are independent, much relevant information is lost. There is strong mutual dependence of one object on the others for properties and meaning. For instance, the image of the sky is always above the imaged hill, assuming that the camera is in an upright position. Hence, if a hill has been determined, then an area strictly below it must not be sky. It may be a lake if it is blue or a road if it is colorless.

The problem with using mutual dependence between different domains interpretations is the overwhelming number of relations that exist. In a set of  $N$  domains, there are  $2^N$  possible relations. The next two sections describe two approaches which allow reduction of combinatorics and, hence, allow a practical use of the mutual dependence of interpretations of domains in a segmented image to get a more reliable global interpretation.

## Model A: "First Order" Structure

As was mentioned above, what is required of a practical system to interpret involved structures of objects is a way of reducing the number of relations involved. This calls for making special simplifying assumptions on the structure of the dependence between objects in the definition of the utility function. One such assumption that seems to be a good compromise between the complexity of expression (1) and the oversimplification of expression (2) is to consider, in addition to the individual regions, pairs of adjacent regions, i.e., pairs of regions that have common boundaries. The new utility function which will approximate the global probability for a given interpretation function INT will be the following expression:

$$\prod_{i=1} P\{\text{INT}(R(i)) \mid \text{measurements inside } R(i)\} \\ \times \prod_{B(i, j)} P\{B(i, j) \text{ is between INT}(R(i)) \text{ and INT}(R(j)) \mid \text{features of } R(i), R(j), \text{ and } B(i, j)\} \quad (3)$$

where  $B(i, j)$  is the boundary between  $R(i)$  and  $R(j)$ .

In this expression, the number of product terms we want to maximize simultaneously is  $O(N)$ , assuming that for a given region the average number of regions that have common boundaries with it is constant, independent of the total number of regions in the segmented scene. An alternative approach is to allow binary relations between any pairs of regions, in which case for a scene that is segmented into  $N$  domains, there always will be  $N + N^2$  terms in the expression that need to be maximized. The nice thing about Eq. (3) is that there is a practical way of maximizing it, using integer programming tree search with pruning (see term definitions in Ref. 6, and Appendix 1 for an algorithmic description).

## Model B: Utilizing Procedural Knowledge Through Sequential Classification

In the previous section the assumption is that, at most, binary relations between the domains will be sufficient and all of those relations should be considered. This assumption constrains the possible features used to analyze a scene, since only a single region or a pair of adjacent regions may be observed at a time; moreover, the consideration of all binary relations may be redundant. A much more flexible mechanism will utilize sequential classification (which is, in a sense, procedural knowledge) to make use of whatever combination of features is important for recognition of a given object. The idea is that the analysis of an object will use the structure of other objects only when essential and can go as far in using many of these other domains as is convenient or beneficial. The decisions as to which features of which object will be helpful in finding the real label for an object with high confidence will be very specific to the configuration around the specific object and, hence, may involve properties of many other domains when needed. In this way, even though we allow involvement of many other domains in analyzing a domain, they will be used only as necessary and, hence, will not create prohibitive combinatorics. The mechanism of sequential classification (classification tree) in Refs. 3 and 4 aims at achieving just that. All that is required is to expand the class of features

which are used by the sequential classification tree to allow references to other objects. One such expansion is to allow, when computing  $P\{INT(R(i)) = j | \dots\}$  the use of features of region  $R(k)$  ( $i \neq k$ ) chosen according to relevance to  $R(i)$ . Another expansion will be to ask about labels (if known) of  $R(k)$  when computing  $P\{INT(R(i)) = j\}$ , ( $i \neq k$ ).

In this way, for instance, if a region is suspected to be a "hammer handle," because it was already found to be elongated, and of the proper size, then the regions bordering the narrow far edges may be tested to see if they were already assigned the label "hammer head." If these are not interpreted yet, then check if their shape is suitable for a "hammer head." Depending on the results of these questions, the probability of the elongated region to be labeled a "hammer handle" may increase or decrease as the result of the above interrogation and will lead to different terminal nodes of the sequential classification tree and, hence, different weight will be assigned to different labels. This type of deduction, without the explicit use of classification trees or probabilities, is used in Refs. 2, 5 and 7. Note that even though the sequential classification mechanism allows the use of properties and labels of many regions to analyze a region, it does not allow direct "feedback." That is, if a region is not assigned meaning yet and is with probability 0.2 "shoe" and 0.8 "hammer head," then this fact cannot be used directly to evaluate the probability of different labels of the region suspected to be "hammer handle." Allowing these probabilities to affect each other in this level will result in a relatively complicated "feedback" equations system.

The mutual dependence may be used as described in the assignment algorithm. The label assignment algorithm in this case (Appendix 2) will be slightly more complicated than in the case of first-order relations, but not significantly more. The added complexity results from the need to remember for each region which new assignments of labels to other regions will affect the probability of different interpretation of that region.

The sequential assignment algorithm (Appendix 2) which is used to find the optimal global interpretation for a segmented image is built around the simple equality:

$$P(A_1 \cap \dots \cap A_n) = P(A_n | A_1, A_2, \dots, A_{n-1}) \\ \cdot P(A_{n-1} | A_1, \dots, A_{n-2}) \\ \dots \cdot P(A_2 | A_1) \cdot P(A_1)$$

which always holds independent of the order of  $A_1, \dots, A_n$ . The algorithm maximizes the right hand of the equality, where in our case for a given interpretation  $INT(\cdot)$ , the event  $A_k$  is "region  $R(I)$  is  $INT(R(i))$ ," when  $R(I)$  is the  $k$ -th region to be assigned a label. All probabilities are computed for the features of the segmented image; the order is decided heuristically by the algorithm.  $A_1$  is the first assignment,  $A_2$  is the second, and so on. The computation mechanism for the different  $P(\cdot)$ 's is the sequential classification mechanism, described briefly below. The major part of the assignment algorithm is to order the tests so as to achieve optimal  $INT(\cdot)$  rapidly.

This approach has one major deficiency: since all the probabilities are only approximated, the resulting

approximation for  $P(A_1 \cap \dots \cap A_n)$  will be order-dependent. That is, if

$$P(R_1 \text{ is a chair and } R_2 \text{ is the leg of a chair})$$

is approximated by

$$P(R(1) \text{ is a chair} | R(2) \text{ is a leg of a chair}) \\ \cdot P(R(2) \text{ is a leg of a chair})$$

it may differ from

$$P(R(2) \text{ is a leg of a chair} | R(1) \text{ is a chair}) \\ \cdot P(R(1) \text{ is a chair})$$

which is another approximation of the combined probability.

This possibility is highly undesirable and may be avoided by approximating the probabilities  $P(A_1 \cap \dots \cap A_n)$  by taking the geometrical averages of all sequential approximations. Since there are  $n!$  ways to order  $n$  elements, there are  $n!$  such approximations and the problem is prohibitive combinatorially. Fortunately, in our case, in computing  $P(A_j | \dots)$  only few  $A_k$ 's are of any importance. In model A, for instance, if  $A_i$  is the event  $R(I)$  being  $INT(R(I))$ , then only events relevant to  $A_i$  are  $A_k$ 's where  $R(K)$  is an adjacent region to  $R(I)$ . Hence, if  $A_i$  is dependent only on 4-10 elements, then  $P(A_i | \dots)$  will have no more than  $2^4$  to  $2^{10}$  different values depending on the order of computation. The weighted geometric average of the different cases of  $P(A_i | \cdot)$  will be

$$\overline{P(A_i)} = (P(A_i | S_1))^{w_1} \\ \cdot P(A_i | S_2)^{w_2} \dots P(A_i | S_n)^{w_n} \\ 0 < w_i \quad \sum_i w_i = 1 \quad (4)$$

where  $S_1, \dots, S_n$  are all subsets of possible events affecting  $A_i$ . Now since  $P(A_j)$  is computed via the classification tree,  $A_j$  will affect  $A_i$  only if in some stage questions about the occurrence of  $A_j$  will be asked while computing  $P(A_i)$ .

In computing the weight for the different terms in Eq. (4), we give equal weights to the values returning from the branch where  $A_j$  is assumed to be known before  $A_i$  and to the branch where  $A_j$  is assumed to become known after  $A_i$ . If this approach is taken, only minor modifications of the assignment algorithm are needed to a get state evaluation function where

$$P(A_1 \cap \dots \cap A_n) = \overline{P(A_1)} \cdot \\ \overline{P(A_2)} \cdot \dots \cdot \overline{P(A_n)}$$

and  $\overline{P(A_i)}$  is the weighted geometrical average of the various order-dependent approximations to  $P(A_i)$ .

### The Sequential Classification Mechanism (Classification Tree)

The classification tree is the mechanism that is used to store the various probabilities and provide a convenient mechanism for learning those probabilities. Full description of the approach is beyond the scope of this paper; see Ref. 3 for a detailed description of the various relevant algorithms and Ref. 4 for the theory. The sequential classification tree in the first model provides the mechanism for computing

$$P\{\text{INT}(R(i)) = j \mid \text{features of } R(i)\}$$

and

$$P\{B(i,j) \text{ is between INT}(i) \text{ and INT}(j) \mid \text{features of } R(i), R(j), B(i,j)\}$$

In the second model, the features used to compute  $P(\text{INT}(R(i)) = j)$  may include other regions' features and labels. As a result, the sequential classification tree provides also for locating, for a given object, the objects that, if they should be labeled, will require recomputing the probabilities of different labels of the given object. This will happen if the sequence of questions about the current object  $R(i)$  leads into nodes of the classification tree where there are questions like: "Is the label of  $R(j)$  X?" With three possible answers: "Yes," "No," and "No label assigned yet." ("No label assigned" means: region  $R(j)$  is not assumed to be known yet.) In this case, when  $R(j)$  becomes known, it will affect the probabilities of the different possible labels of  $R(i)$ , as it will lead to a different terminal node in the classification tree.

### Results and Conclusions

The conclusions of this research are as follows:

(1) There is no way to segment images into regions where each region will stand for the projection of one whole object, without the use of some model of the expected scene. Shadows, noise, and light changes, as results of surface orientation, prohibit such segmentation based only on discontinuity of color or light.

(2) Segmentation based on discontinuity of light reaching the camera, when done with sensitive enough parameters, is very useful for data reduction as it generates relatively few (a few hundred) regions of almost uniform reflectivity structure. Each of these areas is part of the total image of an object, and this data reduction eases the shape recognition and the extraction of global features of the complete objects.

(3) Only a few large areas of uniform light structure, like the sky or the large portion of the road, could be recognized with confidence as they stood alone. For most regions, the mutual contextual constraints were essential for valid recognition. This was especially true for small regions like shadows.

(4) Model A, combined with assignment algorithm 1, is an effective way of allowing references to neighboring structure in order to recognize the smaller objects. Model B, which is more general and allows a more flexible deduction mechanism, is in implementation at the present time.

(5) The supervised learning used for the system described in Ref. 3 is an effective learning system, but the ability to handle a large data base of manually interpreted typical examples for automatic improvement of recognition is desirable.

The results of some experiments are included at the end of the paper, in Figs. 1-3. These figures are cathode ray tube displays of digitized pictures; 5 bits are used to display the original, and 1 bit to draw the boundary lines of regions superimposed on the original. The digitized pictures that the computer processes are in color, with 6 bits in each basic filter (18 bits total for each picture point) but only the 5 significant bits are shown in the pictures.

The basic (nonsemantic) segmentation, based only on continuity of lighting, is described in detail in Ref. 3. The resulting segmented image (which is over-segmented into small regions) is passed to the system, using the assignment algorithm described in Appendix 1, and if two adjacent regions are assigned the same label (meaning), they are assumed to be part of the same object and are merged. A few iterations of reducing the number of regions using the assigned interpretation are demonstrated with each sequence of experiments.

In the outdoor scenes, the allowed objects were sky, hill, tree, road, car, and shadow of car. In the left ventricular angiogram, the allowed objects were the in-heart blood volume, the background organs, and the dark frame. The expected properties of each object were learned under human supervision using the system described in Ref. 3.

### Appendix 1 Assignment Algorithm for Model A

#### Definition of Terms

$R(I)$ ,  $I = 1, \dots, N$  is the  $N$  objects (regions, domains) in the segmented scene.

$B(I, K)$  is the common boundary of  $R(I)$  and  $R(K)$  when it exists.

INTER is the set of  $R(I)$ 's which are assigned a label in the current state of the algorithm.

LABEL( $I$ ) is the label currently assigned to  $R(I) \in \text{INTER}$ .

$J = 1, \dots, M$  is the  $M$  possible labels (interpretations) that can be assigned to objects.

$P(I, J) \rightarrow$  the current score achieved by assigning label  $J$  to object  $I$ .

BEST( $I$ ) is such that  $P(I, \text{BEST}(I)) = \text{Max}_J \{P(I, J)\}$ , where  $1 \leq J \leq M$ , and  $J$  is assignable to  $I$  (to allow backup, a label  $J$  is marked unassignable to  $R(I)$  if, in a given state, all resulting combinations of final states where  $I$  is  $J$  were tested).

NBEST( $I$ ) is such that  $P(I, \text{NBEST}(I)) = \text{Max}_J \{P(I, J)\}$ , where  $1 \leq J \leq M$ ,  $J$  assignable to  $I$   
 $J \neq \text{BEST}(I)$

ASSIGN (LEVEL) contains the name of region assigned meaning in that level.

## Procedures

INIT(I) is the procedure which computes P(I, J), for all labels J, using R(I) features. This procedure pushes R(I) through the sequential classification tree for regions to get into the terminal node in the tree as determined by R(I) features (see Ref. 3 for detailed description).

SORT maintains structure allowing access to the object with highest confidence value (defined below) among the objects not interpreted yet (see Ref. 8, pp. 150-152).

## Algorithm Steps

0. INTER =  $\phi$

1. Call INIT (I) I = 1,  $\dots$ , N

2. SEEN = 0

LEVEL = 0

VALUE (LEVEL) = 1

3. Compute for I = 1,  $\dots$ , N

CONFIDENCE(I) = If P(I, NBEST(I))  $\neq$  0

then  $\frac{P(I, \text{BEST}(I))}{P(I, \text{NBEST}(I))}$  else  $\infty$

(if regions are allowed to be "unidentified objects," then  $\infty$  should be replaced by special formulas).

4. Sort R(I) by CONFIDENCE (I).

5. Choose I  $1 \leq I \leq N$  where R(I) has no assigned label yet such that

CONFIDENCE (I) =  $\text{MAX}_K(\text{CONFIDENCE}(K))$   
K

where:  $1 \leq K \leq N$

and R(K)  $\notin$  INTER

6. Mark BEST (I) as unassignable for I in the current level and for deeper levels. (This will make BEST(I) not assignable again to R(I) should the program get back to that point). Put I in INTER. LABEL(I) = BEST(I).

7. VALUE(LEVEL + 1) = VALUE(LEVEL) \* P(I, BEST(I))

LEVEL = LEVEL + 1; ASSIGN (LEVEL) = I

8. If VALUE(LEVEL)  $\leq$  SEEN then go to B1.

9. If all objects are assigned meaning then: SEEN = VALUE(LEVEL). Store current global interpretation as best seen so far. Go to B1.

10. For all I1 that B(I, I1) exists, do:

P(I1, J1) = P(I1, J1) \*

P(B(I, I1) is between LABEL(I) and J1 |  
features of B(I, I1) R(I) R(I1))

11. For all objects that P(I1, J1) was changed, store old values with level so that when undoing the recent assignment, the old values may be restored.

12. Compute BEST(I), NBEST(I) for all I  $\notin$  INTER if there is an uninterpreted object I1 such that VALUE(LEVEL) \* P(I1, BEST(I1))  $\leq$  SEEN, then go to B1.

13. Go to 4.

B1. If LEVEL = 0, then exit program.

B2. Restore all old values which were changed in current level and mark all pairs of I and J marked unassignable in that level as assignable. Undo the assignment in level, remove ASSIGN(LEVEL) from INTER.

B3. LEVEL = LEVEL - 1.

B4. Go to 12.

## Appendix 2

### Assignment Algorithm for Model B

#### Definition of Terms

R(i) I=1,  $\dots$ , N are the N objects (regions, domains) in the segmented scene.

INTER is the set of R(I)'s which are assigned a label in the current state of the algorithm.

LABEL(I) will be the label assigned to R(I) if R(I)  $\in$  INTER.

J = 1,  $\dots$ , M is the M possible labels (interpretation) that can be assigned to objects.

P(I, J)  $\rightarrow$  the current score achieved by assigning label J to object I.

BEST(I) is such that

$$P(I, \text{BEST}(I)) = \text{Max}_J (P(I, J))$$

where  $1 \leq J \leq M$ , J assignable to I in the current state of the algorithm (see comment in Appendix 1).

NBEST(I) is such that

$$P(I, \text{NBEST}(I)) = \text{Max}_J (P(I, J))$$

where  $1 \leq J \leq M$ , J assignable to I, J  $\neq$  BEST(I)

DEP(I) is the set of all objects such that, should an assignment of label to R(I) occur, then for I1  $\in$  DEP(I), P(I1, J) should be recomputed.

EFF(I) is the set such that if I1  $\in$  EFF(I) is assigned meaning, then P(I, J) should be recomputed. That is, going through the classification tree to compute P(I, J) for various J's, a question about the label of I1 was asked.

Note that I2  $\in$  EFF(I1)  $\equiv$  I1  $\in$  DEP(I2).

ASSIGN(LEVEL) contains the name of the regions assigned meaning in that level.

### Procedures

INIT(I): creates P(I, J), EFF(I) and puts I in DEP(I) when necessary. This procedure makes use, if applicable, of already assigned meanings to other objects in previous iterations of the algorithm, when running R(I) through the sequential classification tree.

SORT: maintains structure allowing access to the object with highest confidence factor among the objects not integrated yet (see Ref. 8, pages 150-182).

### Algorithm Steps

0. INTER =  $\phi$   
DEP(I) = EFF(I) =  $\phi$  I = 1, ..., N
1. Call INIT (I) I = 1, ..., N
2. SEEN = 0  
LEVEL = 0  
VALUE(LEVEL) = 1
3. Compute for I = 1, ..., N  
CONFIDENCE(I) = IF P(I, NBEST(I))  $\neq$  0  
then  $\frac{P(I, BEST(I))}{P(I, NBEST(I))}$  else  $\infty$   
(if regions are allowed to be "unidentified objects," then  $\infty$  should be replaced by special formulas).
4. Sort R(I) by CONFIDENCE(I)
5. Choose I  $1 \leq I \leq N$  where R(I) has no assigned label yet such that  
CONFIDENCE (I) = MAX(CONFIDENCE(K))  
K  
where  $1 \leq K \leq N$ , R(K)  $\notin$  INTER
6. Mark BEST(I) as unassignable for I in the current level and for deeper levels. (This will make BEST(I) not assignable again to R(I) should the program get back to that point). Put I in INTER. Set LABEL(I) = BEST(I).
7. VALUE(LEVEL + 1) + VALUE(LEVEL)\*  
P(I, BEST(I))  
LEVEL = LEVEL + 1; ASSIGN(LEVEL) = I
8. If VALUE(LEVEL)  $\leq$  SEEN then go to B1.
9. If all objects are assigned meaning then:  
(a) SEEN = VALUE(LEVEL). Store current global interpretation as best seen so far, and (b) go to B1.

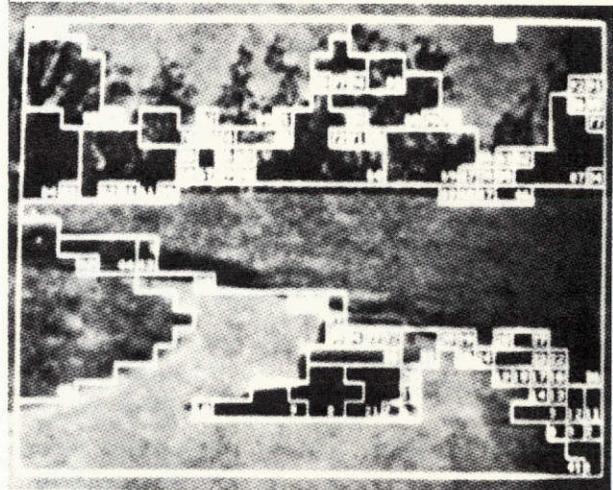
10. Update P(I1, J1) for all I1  $\in$  (DEP(I) - INTERP) by calling INIT(I1) to all those objects.
11. For all objects for which P(I1, J1) was changed, store old values with a level so that when undoing the recent assignment, the old values may be restored.
12. Compute BEST(I), NBEST(I) for all I  $\notin$  INTER if there is an uninterpreted object I1 such that VALUE(LEVEL) \* P(I1, BEST(I1))  $\leq$  SEEN then go to B1.
13. Go to 4.
- B1. If LEVEL = 0, then exit program.
- B2. Restore all old values which were changed in current level and mark all pairs of (I, J) marked as unassignable in that level as assignable. Undo the assignment in level, remove ASSIGN(LEVEL) from INTER.
- B3. LEVEL = LEVEL - 1.
- B4. Go to 12.

### References

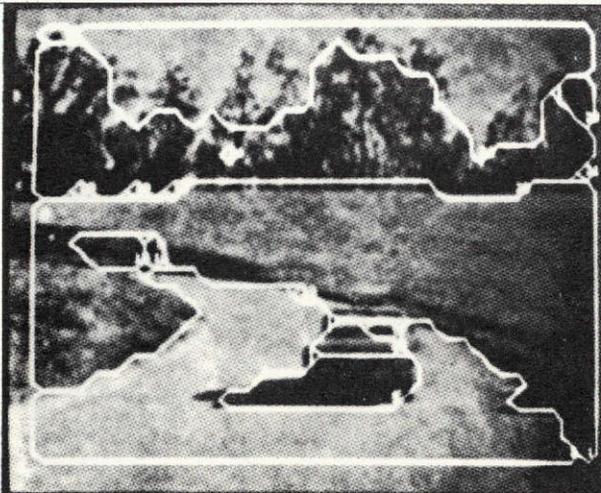
1. Fu, K.S., "Stochastic Grammars and Languages," National Journal on Computer and Information Science, June 1972.
2. Waltz, D.L., Generating Semantic Description from Drawing of Scenes with Shadows, TR-271, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1972.
3. Yakimovsky, Y., Scene Analysis Using Semantic Base for Region Growing, Ph.D. Dissertation, Stanford University, Palo Alto, California, Artificial Intelligence Laboratory Memo AIM-209, June 1973.
4. Fu, K.S., Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, New York, 1968.
5. Lozano-Perez, T., Finding Components on a Circuit Board, Internal Document, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, Sept. 1973.
6. Nilsson, N.J., Problem Solving Methods in Artificial Intelligence, McGraw-Hill, New York, 1971.
7. Freuder, E.C., Suggestions and Advice, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Mass. March 1973.
8. Knuth, D.I., The Art of Computer Programming, Vol. 3, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1973.



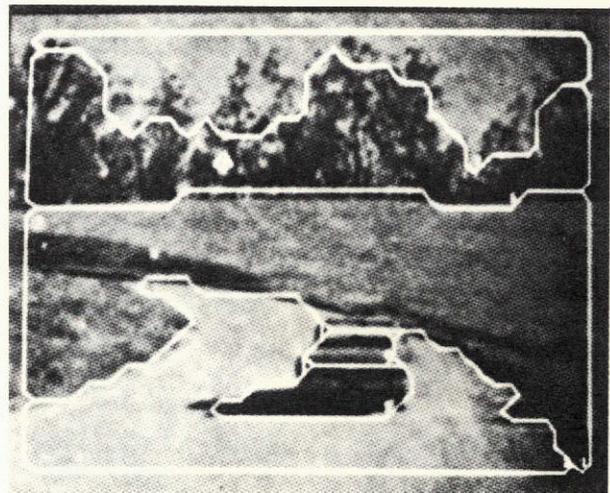
(a) Original picture



(b) Output of the nonsemantic region grower



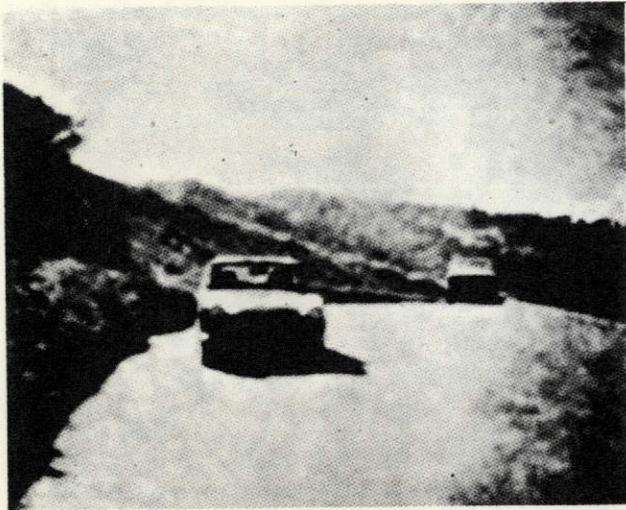
(c) Result of grouping regions by their assigned meaning, taking only regions which were assigned meaning with confidence over 10 to be mergeable.



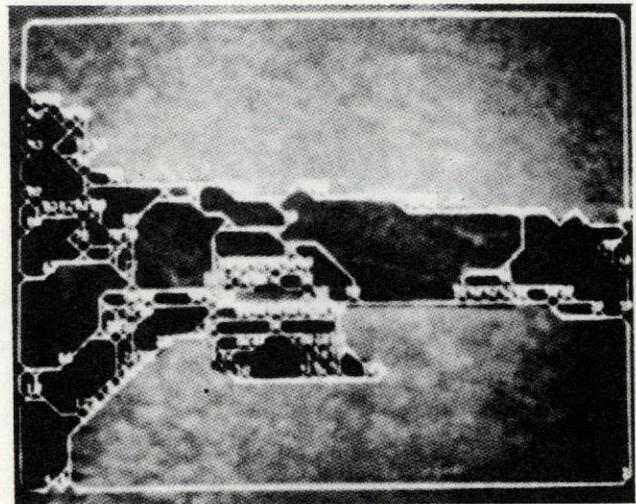
(d) Grouping regions by their assigned meaning, all regions considered mergeable.

Fig. 1. Example of steps of analysis of a road scene

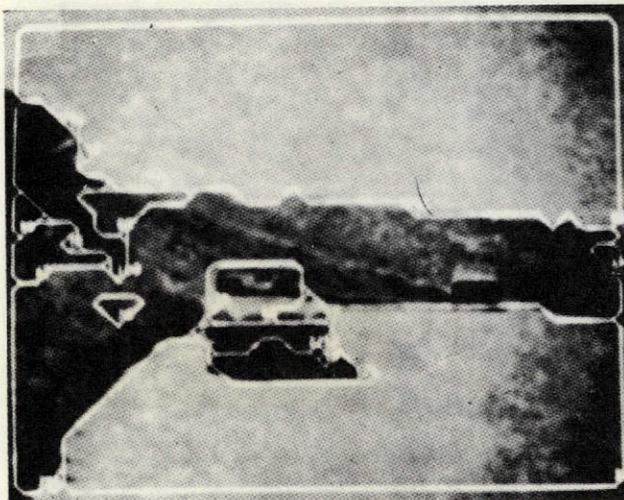
This page is reproduced at the back of the report by a different reproduction method to provide better detail.



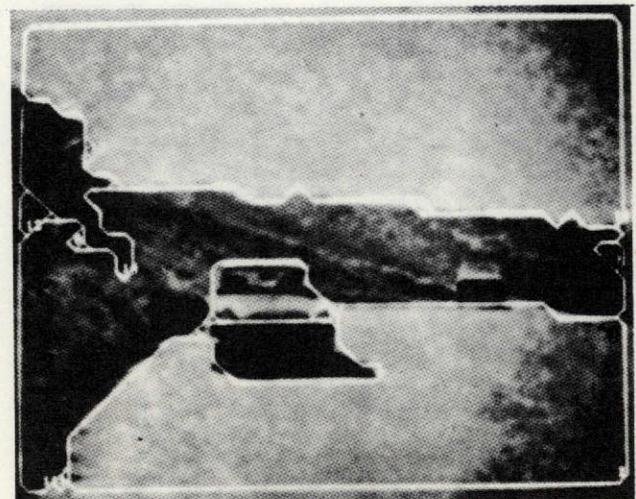
(a) Original picture



(b) Output of the nonsemantic region grower



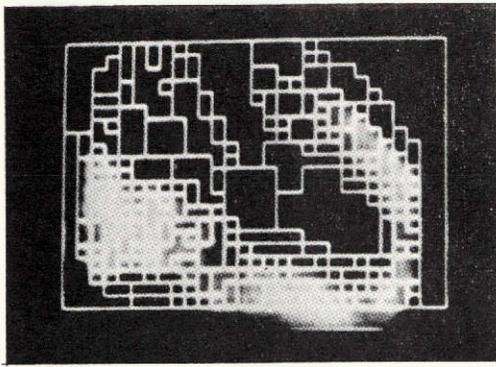
(c) Output of the semantic region grower, pass 1



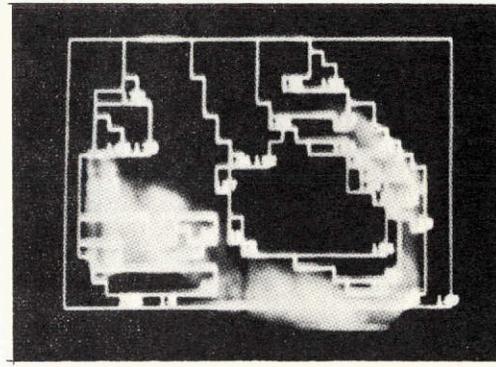
(d) Grouping regions by meaning with confidence 10

Fig. 2. Another example of road scene analysis

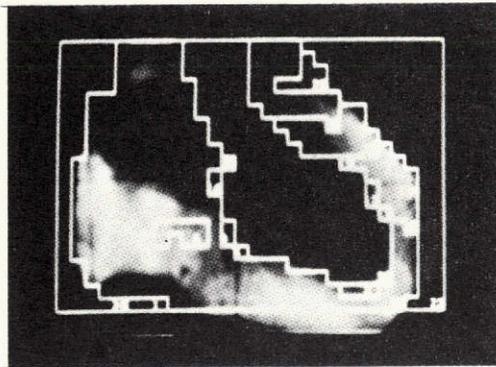
This page is reproduced at the back of the report by a different reproduction method to provide better detail.



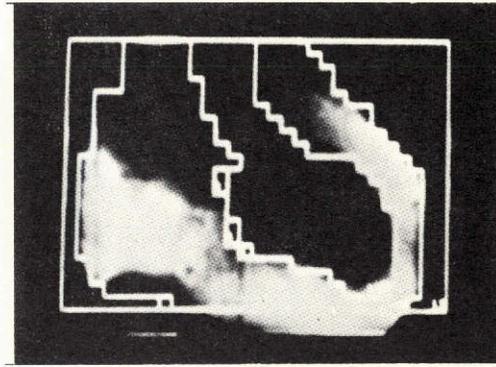
(a) Left ventricular angiogram. Output of the nonsemantic region grower. The stopping criterion is to stop when the merger gets down to 200 regions.



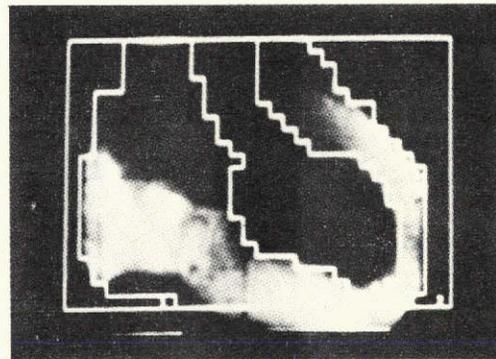
(b) Early iteration of semantic region grower



(c) Result of additional iterations



(d) Result of still more semantic region grower iterations. The region grower used is the grouping of all adjacent regions which are assigned the same meaning by the sequential assignment procedure, before the first assignment with low confidence level occurs. On each iteration, the confidence threshold is lowered.



(e) Final output. The heart interior is the dark center. Around it is the chest cavity and on the two sides there is the dark frame border.

This page is reproduced at the back of the report by a different reproduction method to provide better detail.

Fig. 3. An example of steps of analysis of an X-ray image of a left ventricular angiogram