

Bericht 23

Experiences with Yakimovsky's Algorithm  
for Boundary and Object Detection in Real  
World Images

Hans-Hellmut Nagel

März 1976



### Abstract

Yakimovsky proposed an algorithm for "boundary and object detection in real world images" /1/ and later published /2/ encouraging results obtained with this algorithm on a mini-computer. Applying this algorithm to our raw TV-digitizings yielded results which did not satisfy our requirement of high spatial resolution. Closer analysis of Yakimovsky's algorithm revealed possibilities for improvement which allowed us to obtain better results. Adapting suggestions by Yakimovsky to our situation enabled us to realize further improvements in segmenting raw digitized TV-pictures. Our results are presented and discussed.

-----

We are working on a system to analyze sequences of TV-frames in order to extract descriptions of real world objects - especially moving ones /3/. As input to such a system, traffic on a street intersection has been observed by a commercial monitoring TV-camera and it's output recorded in real-time with an AMPEX analog disk - see Fig. 1a. Sequences of TV-frames are selected under computer control from this analog disk, digitiz-

ed and transferred to a DECSys-10. More details about this setup are given in /3/. Fig. 1b shows a TV-frame reproduced from digitizings with the help of a Thomson-Houston solid state storage tube and Fig. 1c the enlarged section referred to in the following discussion. This section comprises 6144 pixels in 48 lines at 128 pixels each - taking every other TV-line (from one halfpicture) on a section of 96 lines at 128 pixels which allows easy geometry-preserving enlargements without necessity to interpolate grey values. The grey values were digitized with 8 bits and then rounded to five bits to ease their alphanumeric output for diagnostic purposes.

Segmentation of such a section into basic regions of homogeneous grey values and subsequent non-semantic region growing /4/ overtaxed the core available for region-descriptions unless we smoothed the digitizings. Moreover, an iterative weakest-boundary-next region grower can be very time consuming if the basic region finder generates 1000-3000 basic regions. Therefore, Yakimovsky's algorithm /1,2/ has been implemented in the hope to obtain a more manageable segmentation. In the course of verifying this program a difficulty was encountered which could only be overcome by modifying the edge accepting part of Yakimovsky's algorithm.

Yakimovsky uses the concept of an "edge element" between two neighboring raster points. He estimates the probability that the observed grey values in two disjunct environments bordering an edge element can be described by two different Gaussian greylevel distributions and compares this with the probability for explaining the observed grey values by only one Gaussian distribution. From these probabilities he derives a confidence value indicating whether an edge element should be allocated between these neighborhoods or not. Raster points are then aggregated into regions on the criterion that all raster points in a region can be reached from each other by a 4-connected path of neighboring raster points without passing over a local edge confidence maximum or "edge confidence ridge".

In order to discuss the difficulty mentioned, this algorithm must be presented. By observing two logical identities a new form can be found, however, which seems more concise than the original form chosen by Yakimovsky who emphasized the underlying philosophy.

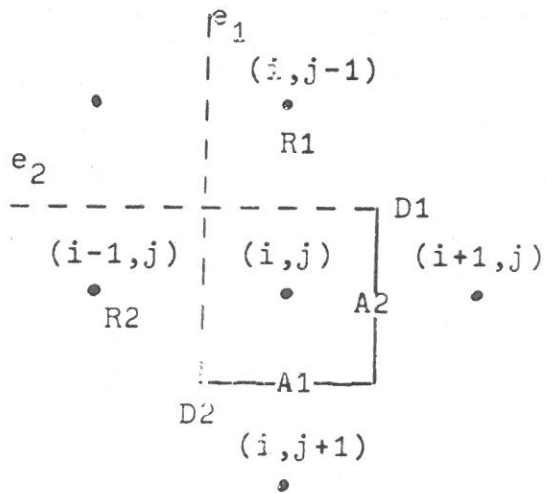


Figure 2

reproduces the notation introduced by Yakimovsky in his Fig. 8 of /1, 2/. D1 becomes true if it is decided to put NO edge unit between  $(i,j-1)$  and  $(i,j)$ . Likewise D2 becomes true if it is decided to put NO edge unit between  $(i-1,j)$  and  $(i,j)$ .

The thresholded edge confidence  $d((i,j),(i-1,j))$  is abbreviated in this figure by A2. Please note that Yakimovsky introduces the thresholded edge confidence between raster point  $(i,j)$  and it's neighbor  $(i+1,j)$  to the right as the distance  $d((i,j),(i-1,j))$  between the raster point  $(i,j)$  and the one to the left of it, namely  $(i-1,j)$ . Likewise the thresholded edge confidence A1 between the raster point  $(i,j)$  and the one  $(i,j+1)$  below it is identified by Yakimovsky with the distance  $d((i,j),(i,j-1))$  between the raster point  $(i,j)$  and the one above it, namely  $(i,j-1)$ . Yakimovsky now defines

$$GD1 \equiv GOOD\_DOWN_1 := (A1 \leq VAL(i,j-1)) \wedge (VAL(i,j-1) \leq VAL(R1)) \quad (1)$$

$$BD1 \equiv BAD\_DOWN_1 := (A1 < VAL(i,j-1)) \wedge (VAL(i,j-1) > VAL(R1)) \quad (2)$$

$$GD2 \equiv GOOD\_DOWN_2 := (A2 \leq VAL(i-1,j)) \wedge (VAL(i-1,j) \leq VAL(R2)) \quad (3)$$

$$BD2 \equiv BAD\_DOWN_2 := (A2 < VAL(i-1,j)) \wedge (VAL(i-1,j) > VAL(R2)) \quad (4)$$

The logical variables  $UP_1$  and  $UP_2$  can be eliminated by observing the following two identities

$$(GOOD\_DOWN \vee UP) \equiv NOT\ BAD\_DOWN \quad (5)$$

$$(NOT\ GOOD\_DOWN \wedge NOT\ BAD\_DOWN) \equiv UP \quad (6)$$

With the abbreviations introduced in equations (1) through (4) and using equations (5, 6) to shift some expressions, the edge accepting part of Yakimovsky's algorithm can be reformulated as:

```

IF GD1
THEN BEGIN
  D1 := TRUE;
  IF GD2
  THEN D2 := TRUE ELSE IF BD2 THEN D2 := FALSE
                                ELSE {i.e. UP2 = TRUE}
                                       D2 := A1 ≥ A2
  END
ELSE IF BD1
  THEN BEGIN D1 := FALSE; D2 := NOT BD2 END
  ELSE {from now on UP1 = TRUE}
    IF GD2
    THEN BEGIN D2 := TRUE; D1 := A2 ≥ A1 END
    ELSE IF BD2
      THEN BEGIN D2 := FALSE; D1 := TRUE END
      ELSE {from now on UP2 = TRUE}
        IF R1 = R2
        THEN BEGIN D1 := TRUE; D2 := TRUE END
        ELSE BEGIN
          D1 := A2 ≥ A1
          D2 := NOT D1
        END;

```

Some meditation - one can't get something for nothing - will convince the reader that the elegance of this formulation reflects the basic simplicity of the decision procedure:

- If GOOD\_DOWN is true, there will be no edge unit, i.e. D will be set true.
- If BAD\_DOWN is true, an edge unit will be allocated, i.e. D will be set false.

If neither of these two conditions is true, one is not forced to either deny an edge unit or allocate it. Two additional considerations now influence the decision:

- If one edge unit is definitely denied, e.g. D1 is set to true, the other edge unit is allocated only if the edge confidence A2 is greater than the edge confidence A1 (and vice versa in the case that D2 is true).
- If the situation is undecided yet for both directions, i.e. to the left and towards above, then no edge unit is allocated if the region R2 bordering the raster point (i,j) in question from the left is identical with the region R1 bordering this point from above. If these regions are different, however, then exactly one edge unit is allocated - consequently towards that neighboring raster point which is connected with the higher edge confidence.

It now becomes obvious where the first objection may be raised: Allocating edge units between raster point (i,j) and its left or upper neighbor, respectively, is in certain situations decided on by comparing the edge confidence between (i,j) and its lower neighbor (i,j+1) with that between (i,j) and its right neighbor (i+1,j) - rather than comparing the edge confidence between (i,j) and its upper neighbor (i,j-1) with the edge confidence between (i,j) and its left neighbor (i-1,j), i.e. this decision is not taken by comparing those edge confidences that correspond directly to the edge units to be allocated.

The first modification consisted in replacing A1 and A2 in the edge accepting algorithm given above by thresholded edge

confidences corresponding directly to the edge units in question. Since A1 and A2 are required by the definition of GD1, BD1, GD2, and BD2, this modification necessitated retaining the thresholded edge confidence values obtained during processing previous points (working on a DECSystem-10 with 65K user area rather than on a minicomputer like Yakimovsky, the segmentation algorithm has been implemented as a strictly one-pass, left to right approach).

The second objection is connected with the definition of GD1, BD1 etc. Yakimovsky requires to consider only thresholded edge confidence values for determination of VAL(i,j) that have not resulted in allocating an edge unit - the rationale being that the edge confidence is infinity whenever an edge unit is allocated and thus will not yield a smaller value for VAL(i,j). But now one is confronted with the situation that VAL(i,j) must be used in the decision process for allocating an edge unit; on the other hand it is only completely defined after all these decisions have been made. The sequential processing of raster points within a line from left to right and descending the lines downwards when the right window edge is reached has one consequence: when deciding upon the last edge unit around a raster point, namely the lower one, all other edge units have already been decided upon and the contribution of their corresponding thresholded edge confidence to VAL(i,j) has been evaluated. The algorithm arrives at the decision about the last edge unit in such a way that the consistency is preserved. However, when deciding about the edge unit between (i,j) and it's left neighbor (i-1,j), the fourth edge unit - between (i-1,j) and (i-1,j+1) - will only be treated a row later and thus VAL(i-1,j) is neither defined nor can it be in the course of the decision about placing an edge unit between (i-1,j) and (i,j).

The second modification consisted, therefore, in replacing VAL(i,j-1) in the definition of GD1 and BD1 by the thresholded



edge confidence between raster points  $(i,j)$  and  $(i,j-1)$  and in replacing  $VAL(i-1,j)$  by the thresholded edge confidence between raster point  $(i,j)$  and  $(i-1,j)$  in the definition of GD2 and BD2.  $VAL(k,l)$  is now only used to determine  $VAL(R)$ .

Figure 3a presents the segmentation of Figure 1c obtained with the original edge accepting part of Yakimovsky and Figure 3b the result of the modified edge acceptor.

Earlier versions of our modified segmentation algorithm applied the single environment depicted in Fig. 4a to determine the edge confidence  $A_1$ . Systematic variation of the edge confidence threshold  $T$  introduced by Yakimovsky did not show the existence of an acceptable threshold yielding segmentations of our raw data with satisfactory spatial resolution. In his original proposal /1/ Yakimovsky hinted at using several environments and picking the one yielding the maximum edge confidence. I applied this suggestion using the three environments depicted in Fig. 4b through 4d. The results are remarkable as can be seen by comparing Fig. 3b with Fig. 3c. One is capable to increase the edge confidence threshold significantly without losing important edges. Since the number of regions obtained using this approach with a threshold parameter  $T = 100$  on equal sections from subsequent TV-frames still varied between 250 and 410, an iterative, most-similar-neighbor-first region merging algorithm as suggested by Yakimovsky /1, 2/ has been applied to the output of the region finder part. With a dissimilarity threshold  $t = 5$  for

$$\ln \frac{v_o^{m+n}}{v_1^m \cdot v_2^n} < t \quad (\text{see /2/})$$

the final segmentation of Fig. 3d has been obtained.

A merging procedure with this threshold  $t$  reduced the number of remaining regions from an average of 340 (taken over 7 sections from subsequent TV-frames) down to 168 or roughly 50 %.

To indicate the extent to which the segmentation result depends upon the two threshold parameters, Figures 5a-d present the scene section from Fig. 1c with a variation of one order of magnitude for the region-merging threshold  $t$ . Although the threshold  $t$  applies to the logarithm of a ratio, a variation of one order of magnitude nevertheless preserves the coarse structure that is important for a semantic interpretation of these segmentations (consult caption for Figures 5a-d).

Figures 6a-d show the same section segmented now varying the edge confidence threshold between  $T = 10$  and 2000 at a constant region-merging dissimilarity threshold  $t = 5$ . These figures demonstrate that the "semantics" of the picture is retained by the segmentation over a rather wide choice of threshold parameter values. Figure 7a depicts a TV-picture reproduced from the segmentation of Fig. 5b and Fig. 7b presents a subsection from Fig. 7a enlarged. Considering the fact that Fig. 7b has been obtained by applying the modified Yakimovsky algorithm to a minuscule subsection from the original raw TV-digitizings, the spatial resolution of this segmentation algorithm now seems acceptable for our purposes.

There is one further observation in connection with Yakimovsky's algorithm which I want to report although it has not yet been possible to find a satisfying solution. Looking at the output of the edge acceptor one frequently finds series of short - i.e. one raster element long - parallel cracks or a "comb". Closer analysis shows that such a "comb" is due to the sequential nature of the edge accepting algorithm combined with a special constellation of edge confidence as a function of raster point location relative to the processing direction. Consider the combination of thresholded edge confidences as depicted in Figure 8. There are two important points to note: The persistent decrease of edge confidence when going from left to right in a row ( $j = \text{constant}$ ) and the fact that the

thresholded edge confidence towards the right neighbor (larger  $i$ ) is greater than the one towards the upper neighbor (smaller  $j$ ).

To understand the comb-effect let us assume a (vertical) region boundary between  $i-2$  and  $i-1$  for all  $j$  values depicted. First consider the point  $(i-1, j+1)$ . Since the edge confidence increases with increasing  $j$ , no edge unit is allocated between  $(i-1, j)$  and  $(i-1, j+1)$ . Since an edge unit has been allocated between  $(i-2, j+1)$  and  $(i-1, j+1)$ ,  $VAL(i-1, j+1)$  is set equal to the edge confidence between  $(i-1, j)$  and  $(i-1, j+1)$ . Proceeding to raster point  $(i, j+1)$ , it can be seen that the edge confidence  $A2$  between  $i-1$  and  $i$  is both larger than  $A2$  between  $i$  and  $i+1$  as well as being larger than  $VAL(i-1, j+1)$ , hence  $BAD\_DOWN_2$  becomes true for raster point  $(i, j+1)$  and consequently another edge unit is allocated parallel to the one between  $(i-2, j+1)$  and  $(i-1, j+1)$ . Since the edge confidence between raster points  $(i, j)$  and  $(i, j+1)$  still increases, no edge unit will be allocated between these points,  $VAL(i, j+1)$  will be set equal to this edge confidence and the conditions prevailing at raster point  $(i-1, j+1)$  are reproduced for its right neighbor  $(i, j+1)$  - and so on: a series of parallel cracks will be generated as long as this constellation of thresholded edge confidence values prevails. The reader may convince himself that the inverse slope of thresholded edge confidence values with respect to the processing direction will not result in such a "comb" of cracks, i.e. the comb results from the directional preference of the algorithm due to its sequential nature combined with a "little too much locality" for coping correctly with such situations.

It should be pointed out, however, that this property does not adversely affect the segmenting power of the modified Yakimovsky algorithm. It can simply be coped with by dropping short cracks - at the only cost of increased overhead (generate a description for the starting edge, detect it as a short crack, remove the description and reclaim the core reserved for it).

Suppressing cracks of length 1 removed about 340 out of 2800 edge lines (both boundaries and cracks) for an edge confidence threshold setting of  $T = 100$  in the picture section of Fig. 1c. About half of these 340 very short cracks were due to the comb effect.

### Conclusion

The picture segmentation algorithm proposed in /1/ and discussed by Yakimovsky has been analyzed and slightly modified. It is conjectured that Yakimovsky himself was not interested in extremely high spatial resolution and consequently did not investigate the details discussed in this contribution. The slightly modified algorithm supports Yakimovsky's basic claim by demonstrating a satisfactory segmentation with a high spatial resolution which is only weakly influenced by the specific choice of the two threshold parameters. The resulting segmentation of raw TV-digitizings from an untuned camera/analog-disk sensor-setup does not require any semantic input /5,6,7/ and nevertheless preserves the possibilities for a semantic interpretation to an astounding degree. This observation tends to support the hypothesis of Marr /8/ that a proper non-semantic analysis may enable one to extract much more information from the digitized greyscale pictures than might have been suspected during certain phases of scene-analysis history. The modified Yakimovsky algorithm seems to be a reliable base for our continuing work on the analysis of sequences of TV-scenes.

### Acknowledgements

R. Bertelsmeier, H. Kemen, B. Neumann, and B. Radig with the support of P. Cord and I. Heer have collaborated in the design and implementation of our picture laboratory setup. During the preparation of this contribution I enjoyed several discussions with B. Neumann - who in addition kindly helped me to prepare the photographs - and B. Radig. Originally, I started by supervising an undergraduate term project to compare Yakimovsky's algorithm with existing ones in our laboratory. Among my students especially R. Kraasch and W. Zach are to be mentioned for their ideas and efforts to implement this algorithm at our installation. It is a pleasure to thank Mrs. Jancke for her help in preparing this report.

I am grateful to the Federal Ministry for Research and Technology and the Freie und Hansestadt Hamburg for their financial support extended through the university which enabled our group to build up a picture laboratory.

References

- /1/ Y. Yakimovsky      Boundary and Object Detection in Real World Images  
 Technical Memorandum 33-709  
 Jet Propulsion Laboratory, Cal. Institute of Technology, Pasadena, Cal., November 15, 1974
- /2/ Y. Yakimovsky      Boundary and Object Detection in Real World Images  
 Advance Papers of the Fourth International Joint Conference on Artificial Intelligence, p. 695  
 Tbilisi, Georgia, USSR, 3-8 September 1975
- /3/ H.-H. Nagel          Towards Analyzing Sequences of Scenes Containing Objects in Motion  
 Contribution to the 2nd meeting GI-Fachgruppe "Artificial Intelligence", 7 October 1975, University of Dortmund, Berichte des Instituts fuer Informatik Dortmund (G. Veenker, ed.), Nr. 13/75, p. 126
- /4/ see, e.g., R.O. Duda and P. Hart, Pattern Classification and Scene Analysis  
 J. Wiley & Sons, New York 1973  
 or the relevant articles quoted in /1/ through /3/.
- /5/ Y. Yakimovsky      Scene Analysis Using a Semantics Base for Region Growing  
 STAN-CS-73-380 Stanford University, June 1973
- /6/ Y. Yakimovsky      On the Recognition of Complex Structures  
 Proc. 2nd International Joint Conference on Pattern Recognition, p. 345  
 Copenhagen, August 13-15, 1974
- /7/ J.M. Tenenbaum and S. Weyl, A Region-Analysis Subsystem for Interactive Scene Analysis  
 Fourth International Joint Conference on Artificial Intelligence, p. 682  
 Tbilisi, Georgia, USSR, 3-8 September, 1975
- /8/ D. Marr              Early Processing of Visual Information  
 AI Memo 340, Massachusetts Institute of Technology, December 1975

Figure Captions

- Fig. 1a Street intersection reproduced from AMPEX TV-analog disk
- Fig. 1b Same scene reproduced from TV-digitizings by means of a Thomson-Houston solid state storage tube. The horizontal stripes are due to the linewise transfer of digitizings between a minicomputer and the storage tube.
- Fig. 1c Center section of fig. 1b (48 lines  $\times$  128 pixels 6 bit grey values from even half picture) enlarged 4 times horizontally and 6 times vertically to preserve geometry. The digitizings of this section are used for the following discussion.
- Fig. 2 Notation used in discussion.
- Fig. 3a Segmentation of digitizings from fig. 1c obtained by the original version of Yakimovsky's algorithm. The edge confidence threshold is  $T = 30$ ; only the environment of fig. 4b has been used to produce this segmentation.
- Fig. 3b Segmentation of the same digitizings as in fig. 3a with same edge confidence threshold by the modified algorithm.
- Fig. 3c Conditions as in fig. 3b but picking the maximum edge confidence obtained for all three environments sketched in fig. 4b through 4d.
- Fig. 3d Conditions as in fig. 3c but applying a most-similar-neighbor-first iterative region-merging algorithm to the output of the modified region-finder with dissimilarity threshold  $t = 5$ .
- Fig. 4 Different environments used to determine the edge confidence between raster points.
- Fig. 5 Segmentation of digitizings from fig. 1c using the modified edge acceptor, three environments for picking the maximum edge confidence and iterative most-similar-neighbor-first region-merging as in fig. 3d but with different threshold parameters. The edge confidence threshold has been set to  $T = 100$  and the dissimilarity threshold  $t$  varied between
- $t = 3$  (fig. 5a)
  - $t = 5$  (fig. 5b: only difference to fig. 3d is  $T = 100$  rather than  $T = 30$ !)
  - $t = 10$  (fig. 5c)
  - $t = 30$  (fig. 5d)
- Please note that the coarse structure of the main objects (car from side, car above it, traffic sign to the right) is retained.
- Fig. 6 Variation of the edge confidence threshold  $T$  over more than two orders of magnitude keeping the dissimilarity threshold  $t$  for region-merging constant at  $t = 5$ .

T = 10 (fig. 6a)  
T = 30 (fig. 3d)  
T = 100 (fig. 5b)  
T = 500 (fig. 6b)  
T = 1000 (fig. 6c)  
T = 2000 (fig. 6d)

Please note the message: only at  $T = 2000$  regions with a substantially different interpretation don't remain separated.

Fig. 7a Greyscale reproduction by means of a TV-storage tube from the segmentation of fig. 1c as depicted in fig. 5b.

Fig. 7b Enlargement from fig. 7a around the car in the upper half. Although this car is represented in the raw digitizings by not more than 15 TV lines with 13 pixels each, the dominant features - front window, car top, hood, left and right fender - are preserved by the segmentation in a way that is accessible to a correct semantic interpretation.

Fig. 8 Distribution of thresholded edge confidence values which result in a "comb" of cracks when processed from left to right with small row numbers  $j$  first.

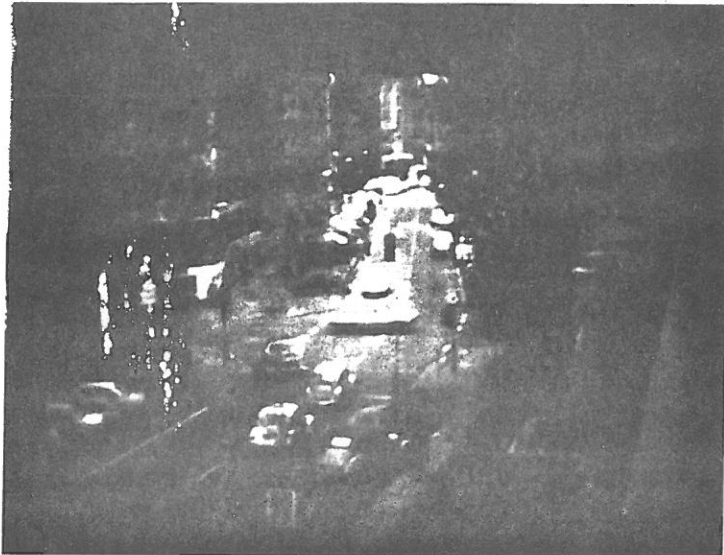


Fig. 1a

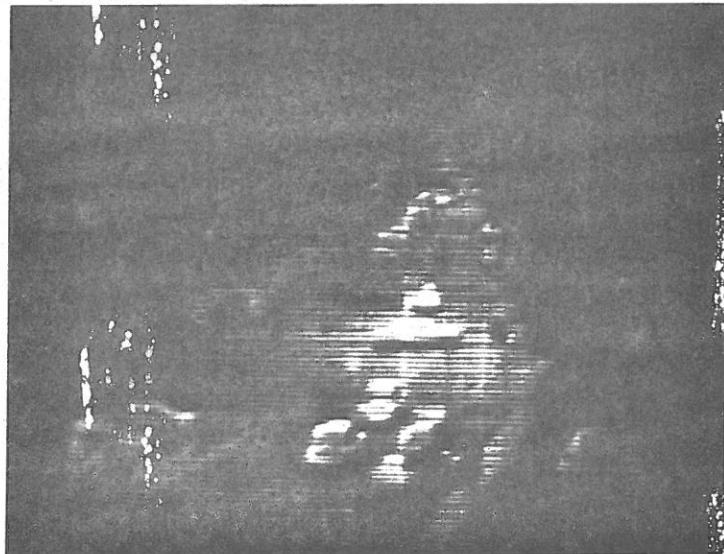


Fig. 1b

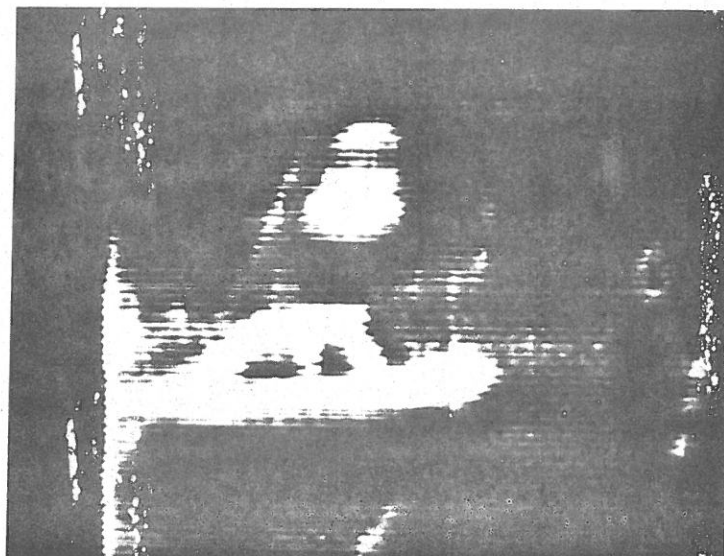


Fig. 1c





Fig. 1a

due to reproduction difficulties  
we enclose another attempt

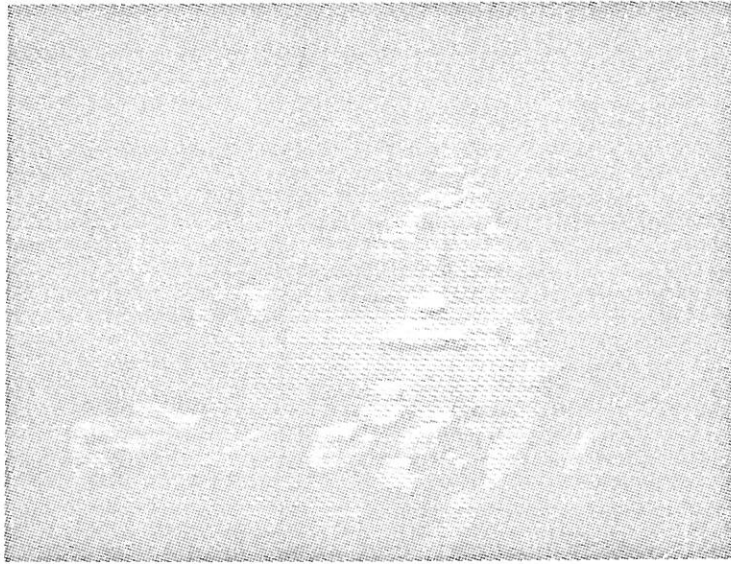


Fig. 1b

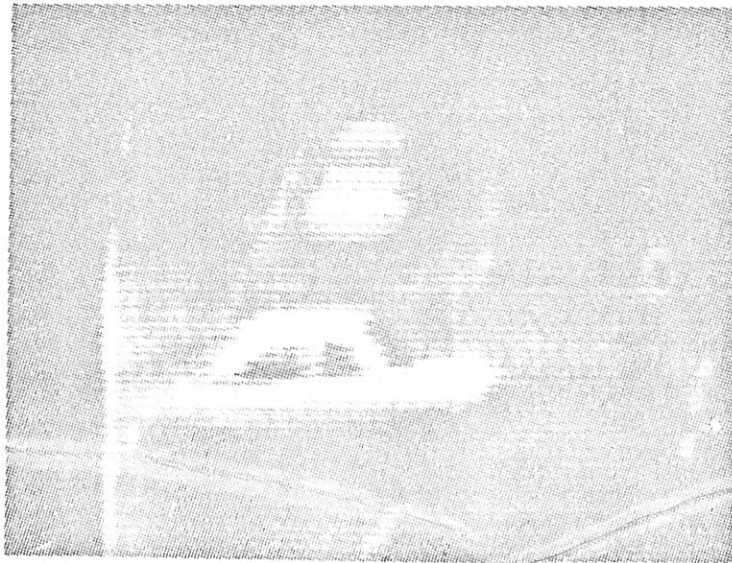


Fig. 1c



VERSION V1C3 AM 17-MAR-76 UM 20138149.488 UHR FUER HALBBILD-AUSSCHNITT AUS DSK 1 BILD ,V86 (000306,000100)  
ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 30,0 UND G-SCHWELLE = 9,0

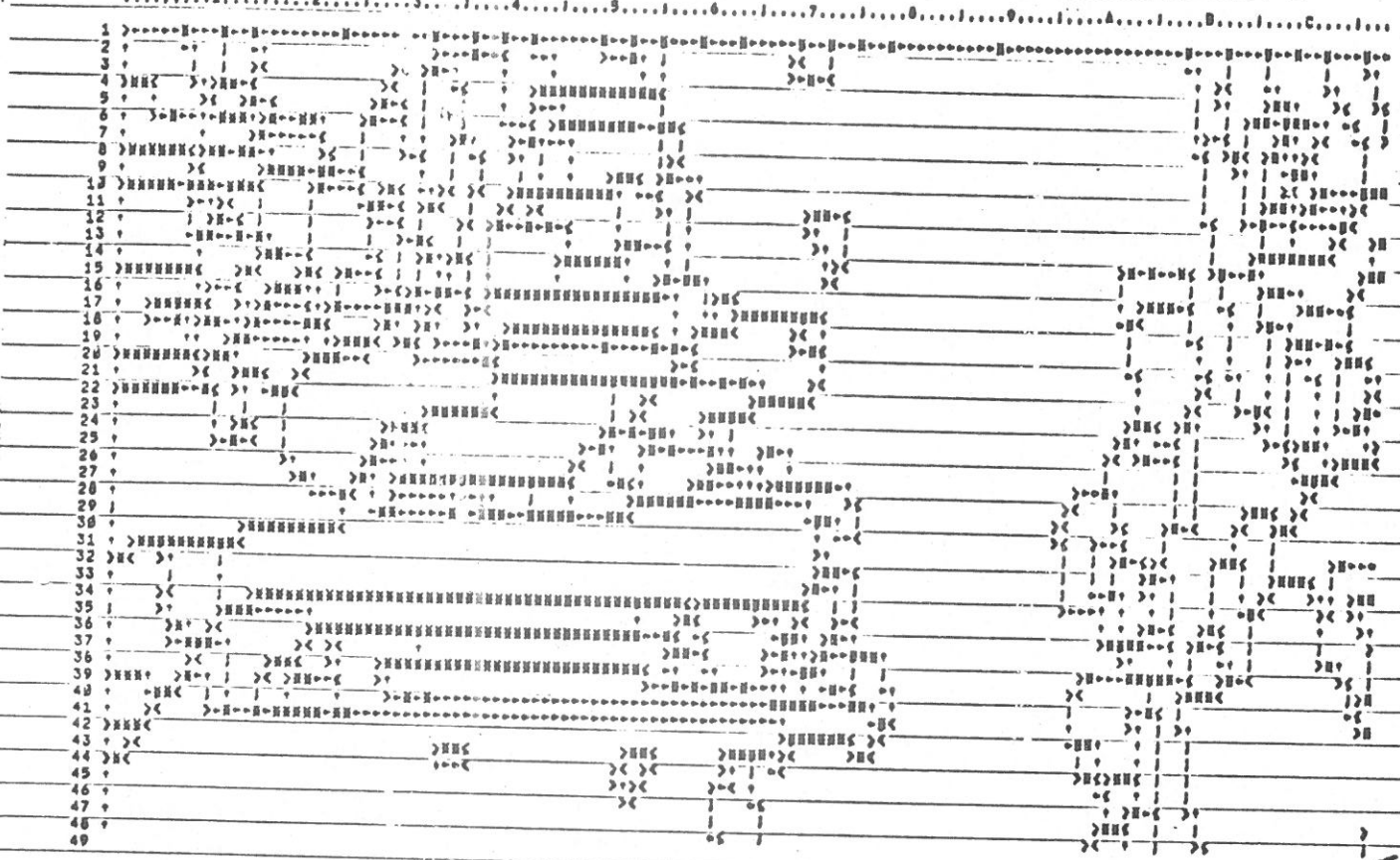


Fig. 3a

VERSION V1C2 AM 17-MAR-76 UM 2149119.568 UHR FUER HALBBILD-AUSSCHNITT AUS DSK 1 BILD ,V86 (000306,000100)  
ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 32,0 UND G-SCHWELLE = 9,0

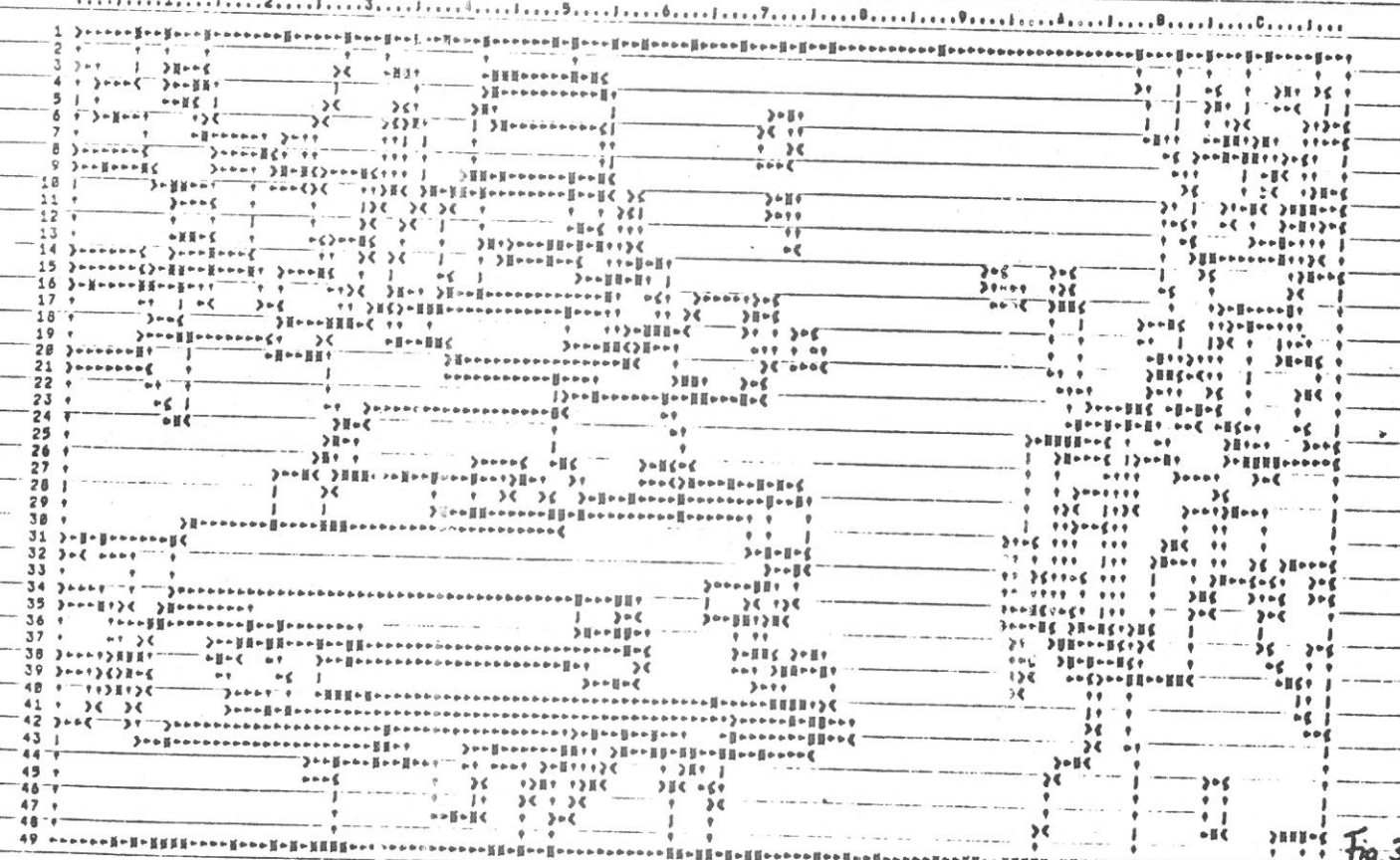


Fig. 3b

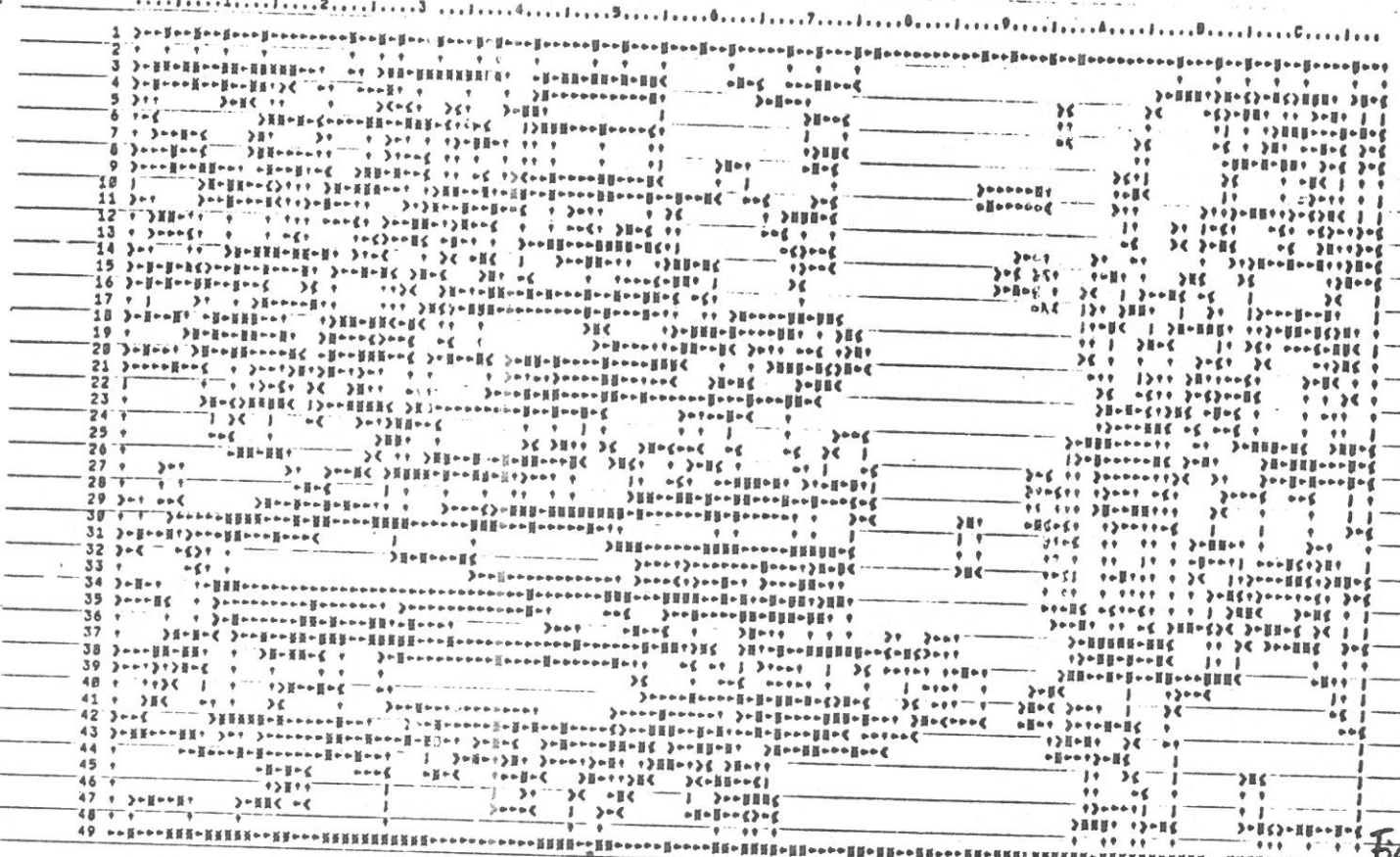


Fig. 3c

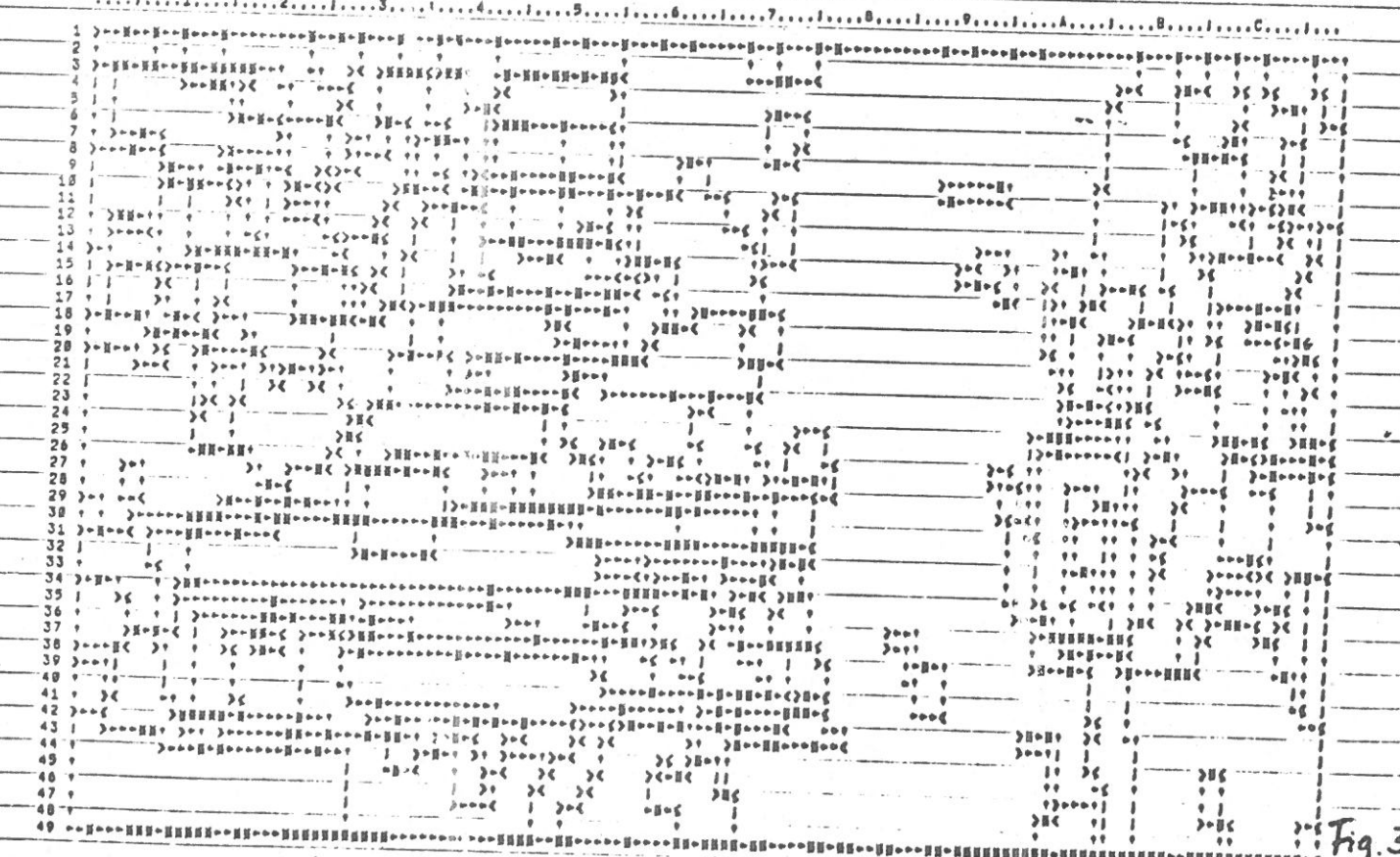
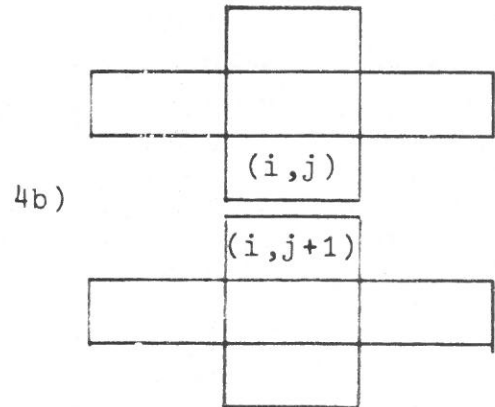
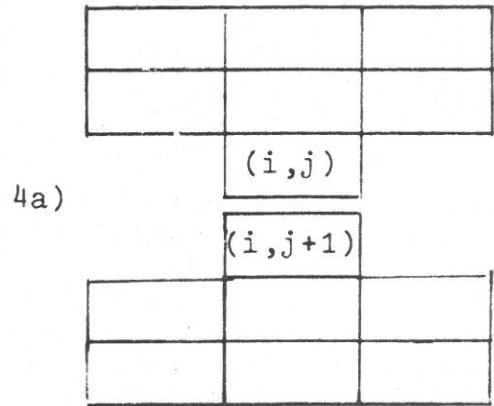


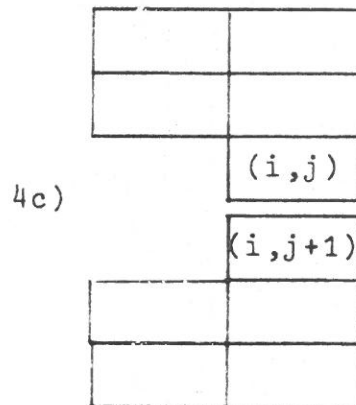
Fig. 3d

Environment to determine the edge confidence between raster point  $(i,j)$  and it's neighbor below,  $(i,j+1)$ . The environment for determination of A2 are obtained by rotating the figures through 90 degree.

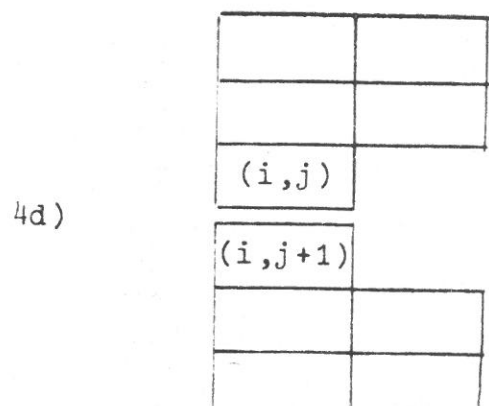
Single environment used in earlier versions of our segmentation program



In later versions of our segmentation program all three environments b through d are evaluated for every point  $(i,j)$  and the edge confidence is chosen to be the maximum value obtained with one of these three environments



The number of raster points in the environment has been reduced from 7 to 5 in order to restrict the maximum extension to 3 lines by 3 columns and nevertheless retain the same area for all three environments in fig. 4b through 4d.



VERSION V3G AM 16-MAR-76 UM 23:41:26,748 UHR FUER HALBBILD-AUSSCHNITT AUS DSK I BILD ,V36 (000306,000100)

ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 100,0 UND G-SCHWELLE = 3,0

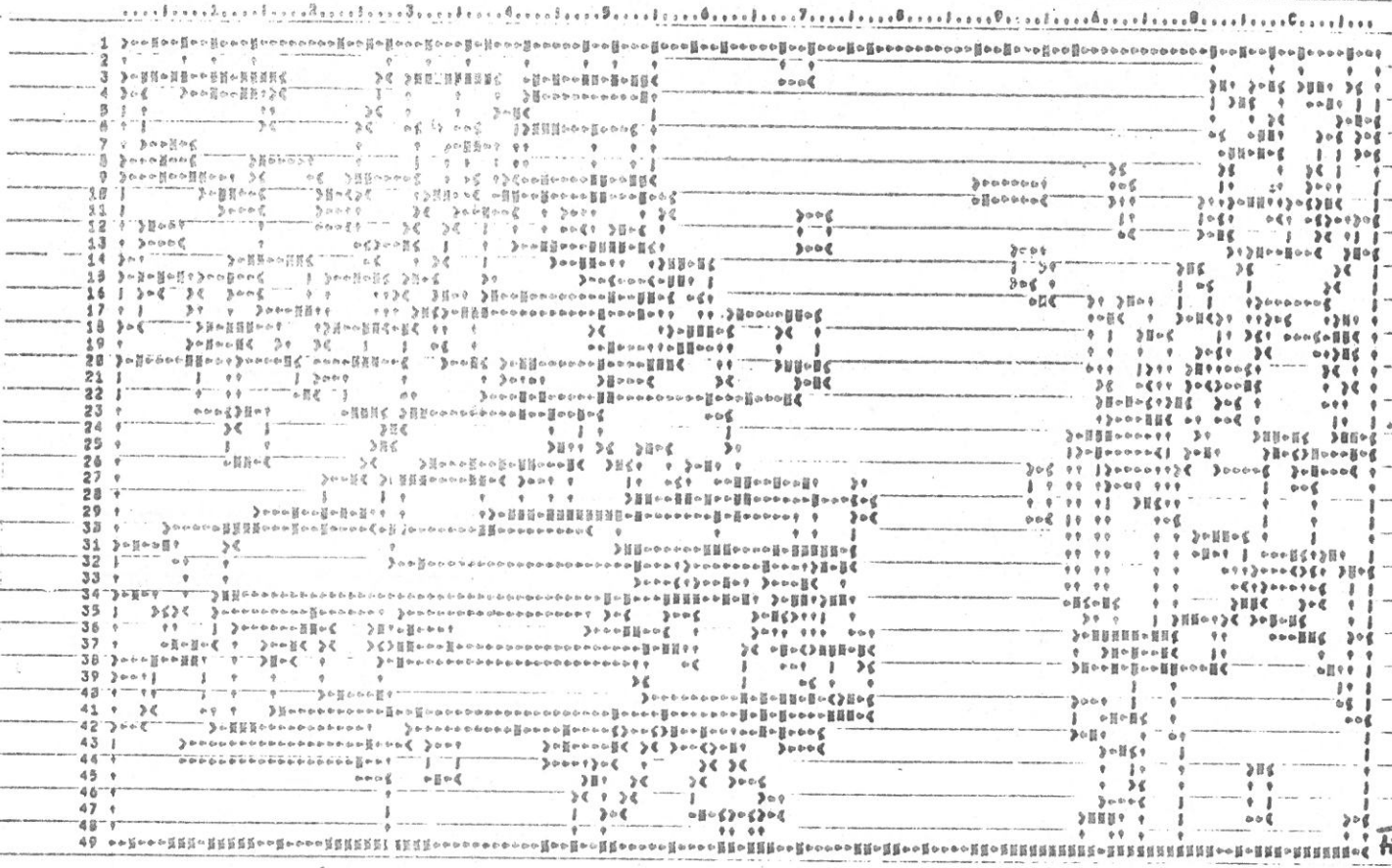


Fig. 5a

VERSION V3G AM 17-MAR-76 UM 0:44:43,838 UHR FUER HALBBILD-AUSSCHNITT AUS DSK I BILD ,V36 (000306,000100)

ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 100,0 UND G-SCHWELLE = 3,0

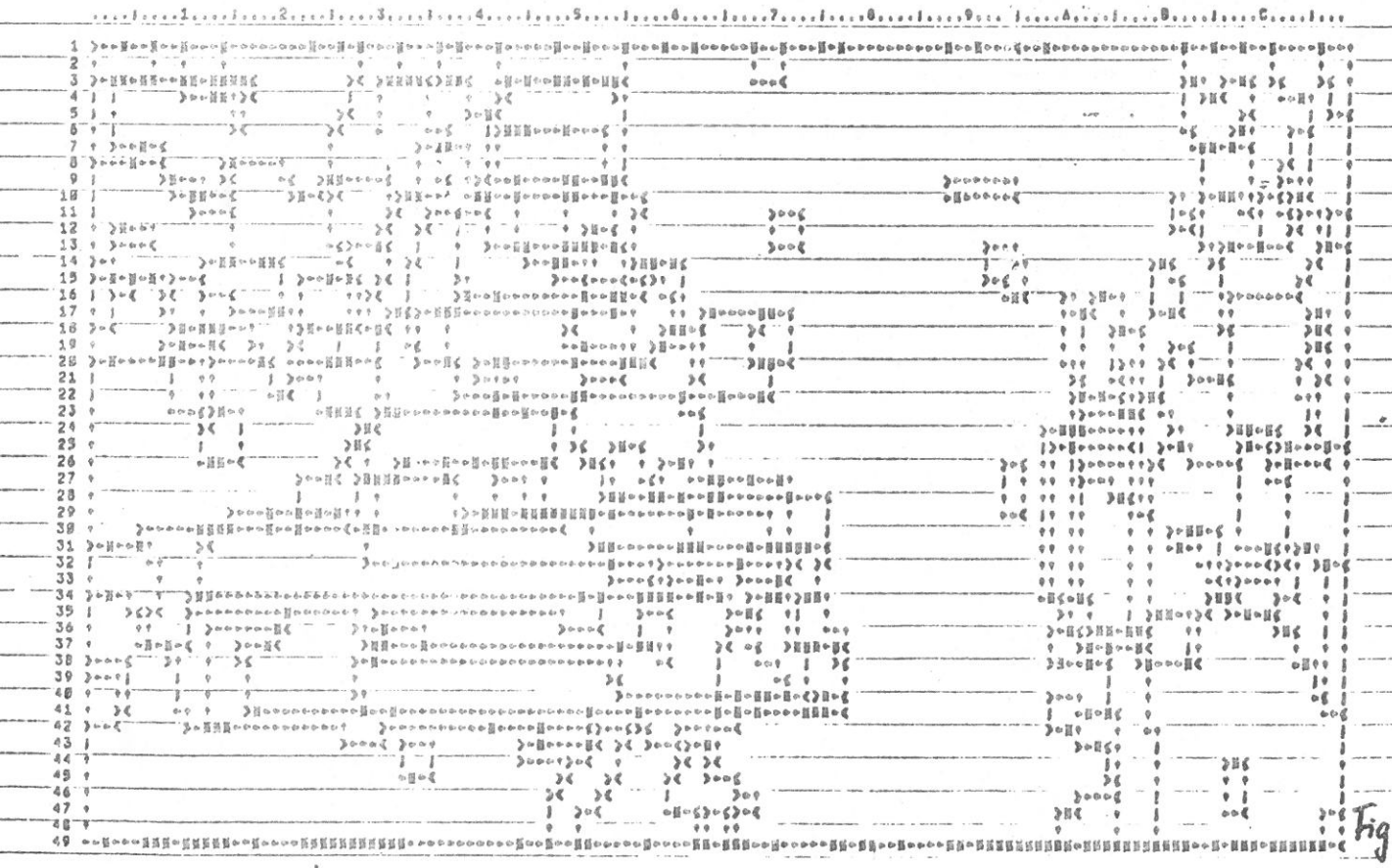


Fig. 5b

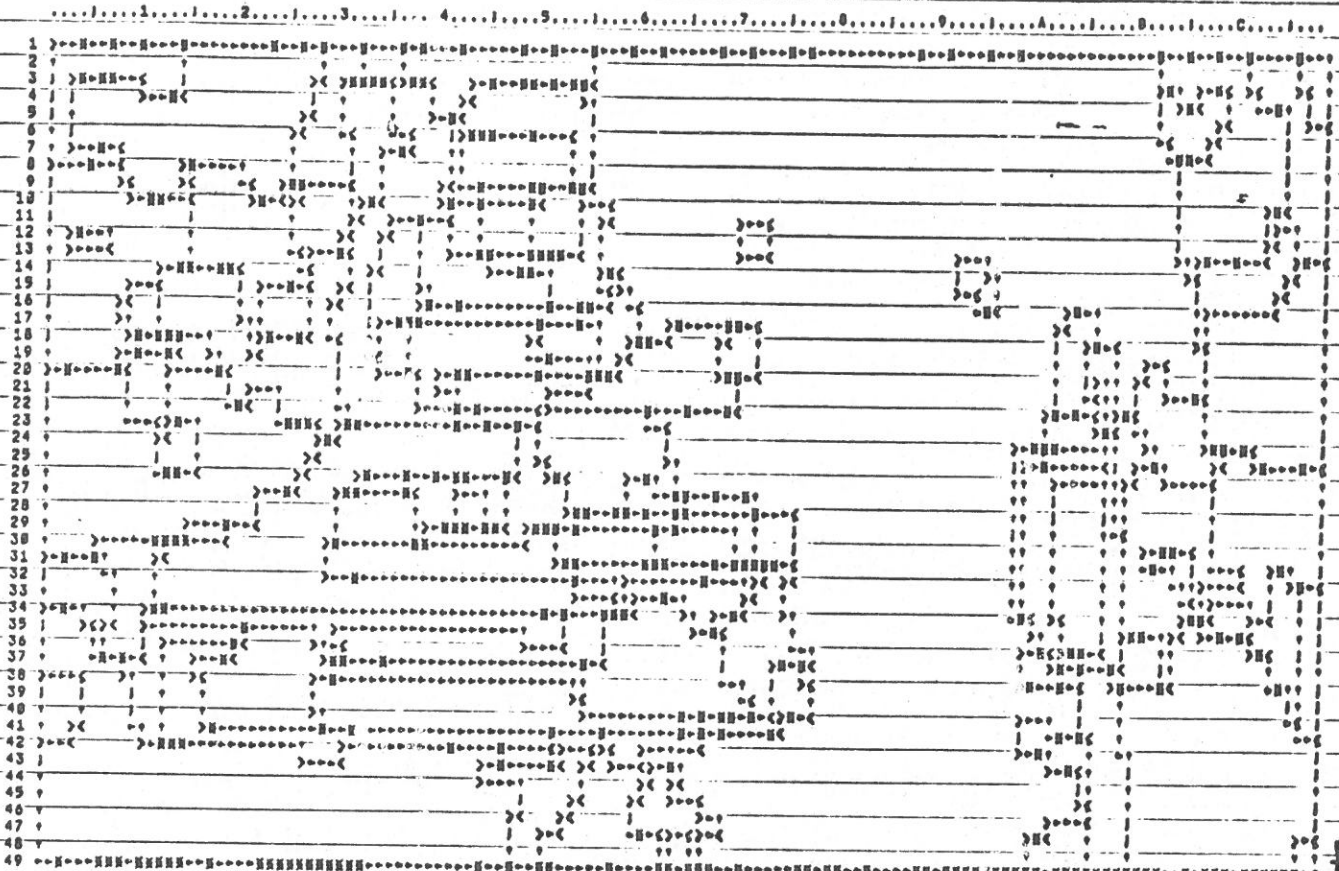


Fig. 5c

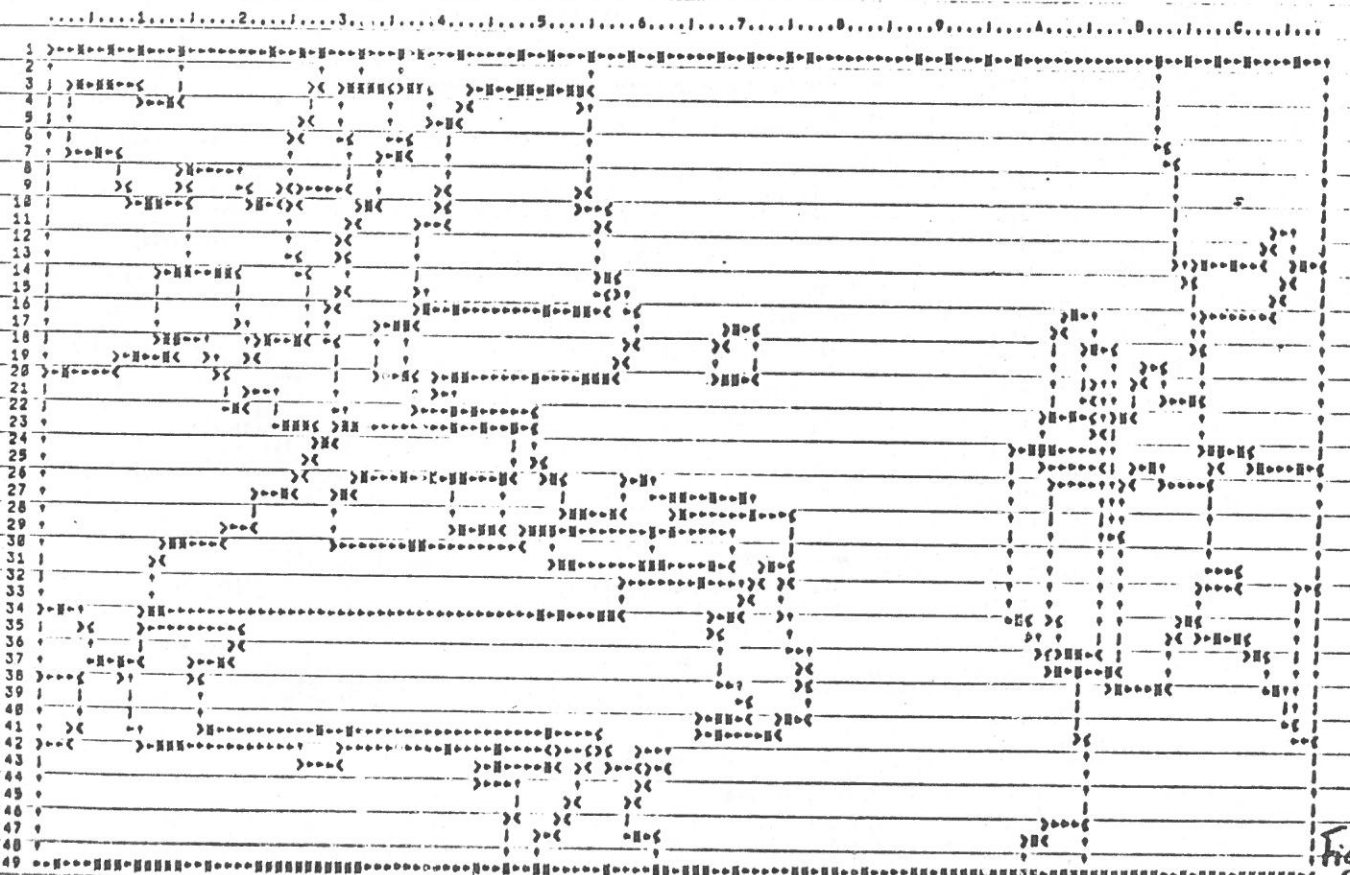


Fig. 5d

VERSION VIC AM 17-MAR-76 UM 11.7134.668 UHR FUER HALBBILD-AUSSCHNITT AUS OSK 1 BILD .V06 (000306,000100)

ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 18,0 UND G-SCHWELLE = 9,0

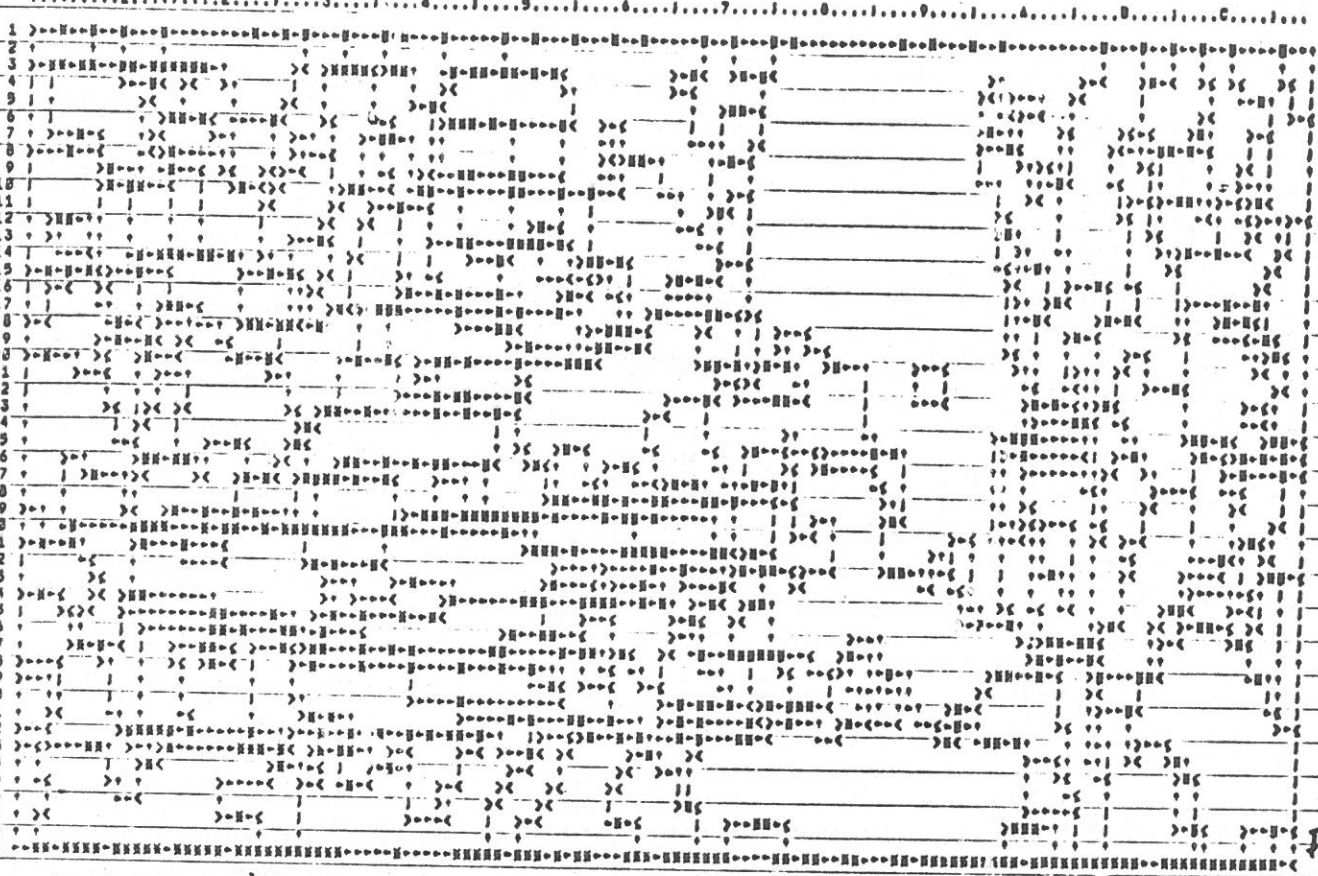


Fig. 6a

VERSION VIC AM 17-MAR-76 UM 11.4219.768 UHR FUER HALBBILD-AUSSCHNITT AUS OSK 1 BILD .V06 (000306,000100)

ERSTE ZEILE = 232 ERSTE SPALTE = 231 MIT THRESHOLD = 500,0 UND G-SCHWELLE = 9,0

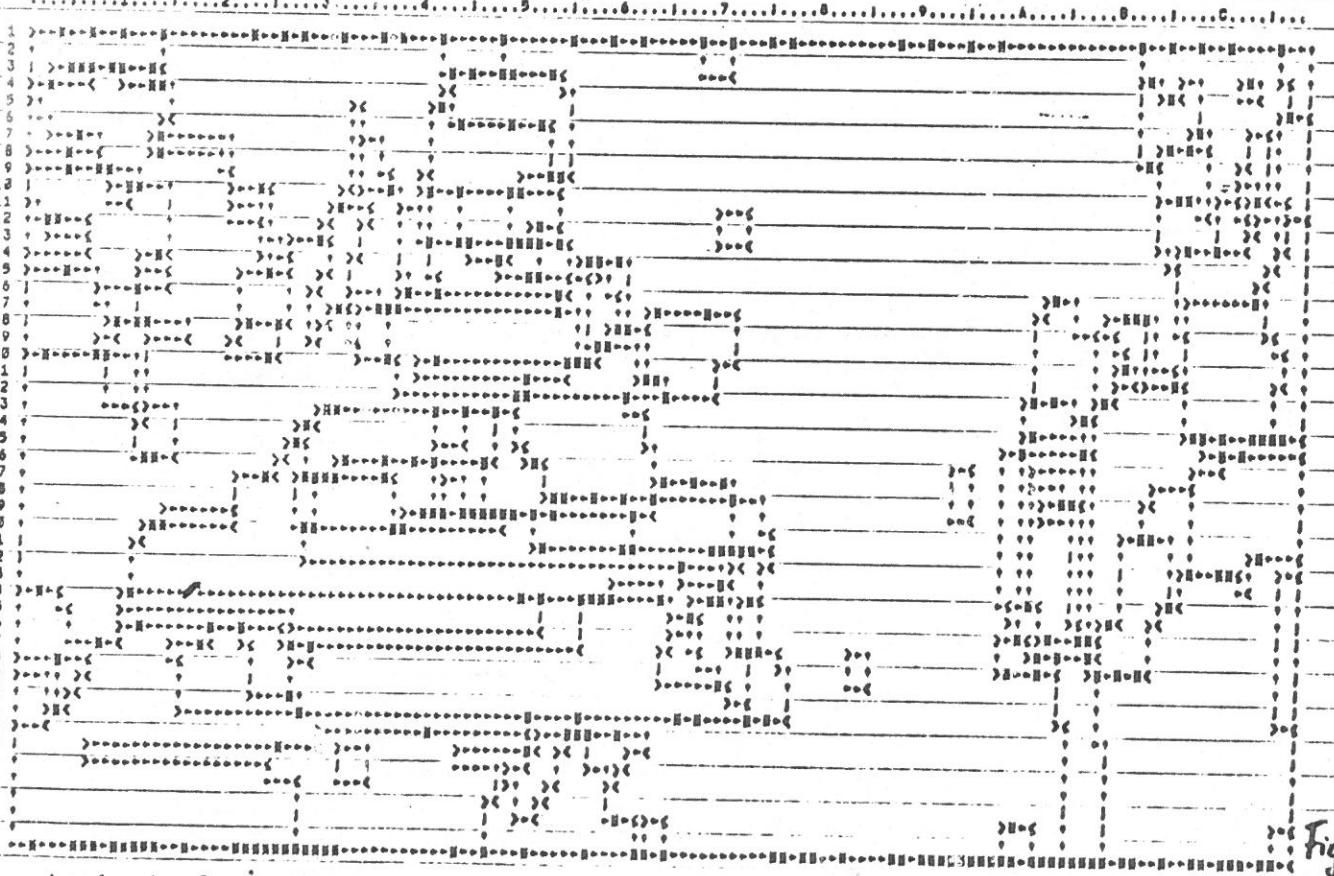


Fig. 6b





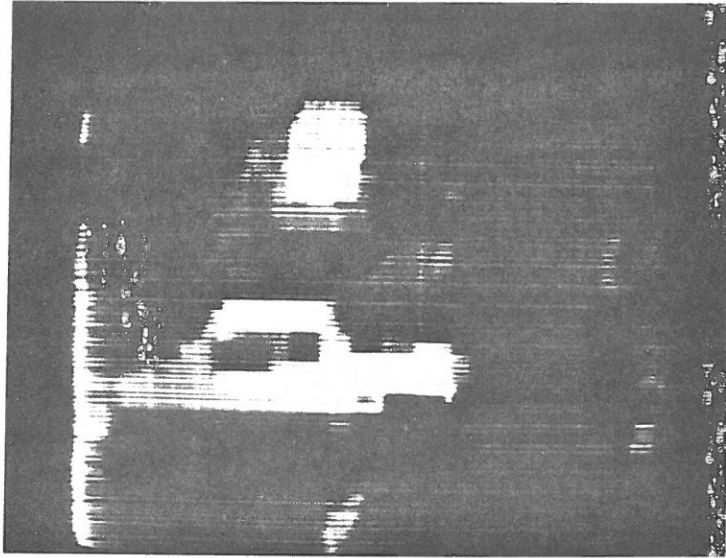


Fig. 7a

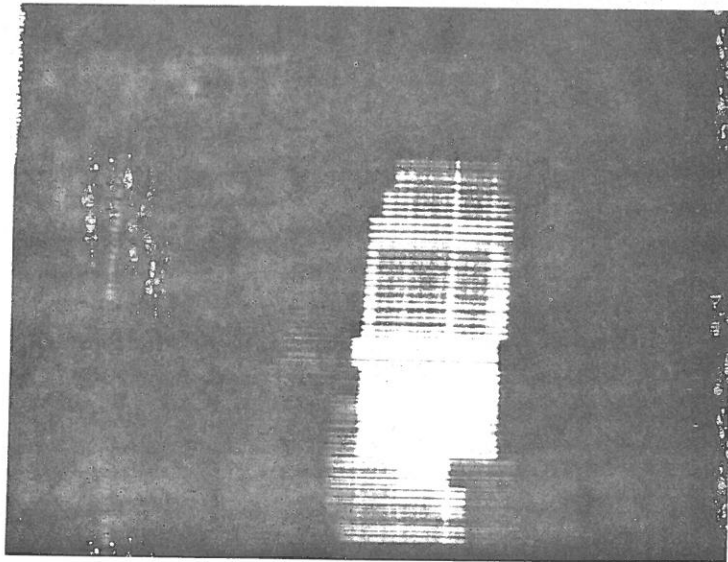


Fig. 7b