

~~040172~~

043072

Final Report for the LARS/Purdue-
IBM Houston Scientific Center
Joint Study Program

by

P. E. Anuta, E. M. Rodd,
R. E. Jensen and P. R. Tobias

The Laboratory for Applications of Remote Sensing

Purdue University
Lafayette, Indiana

Final Report
for the
LARS/Purdue - IBM Houston Scientific Center
Joint Study Program
by
P. E. Anuta¹, E. M. Rodd², R. E. Jensen²
and P. R. Tobias²

I. Summary

This report describes work performed during the second part of the LARS/IBM HSC Joint Study program which began April 1, 1971 and terminated April 1, 1972. Work during the first part of the study ending November 30, 1971 concentrated on analysis of the HSC vidicon film scanning system. Included was a comparison of data derived from this system and the LARS/U. of Michigan airborne multispectral scanner system. Comparison of vidicon digitized film and film digitized on a rotating drum microdensitometer was also included. This work is reported in IBM Publication No. 320.2421 "First Interim Progress Report for IBM Houston Scientific Center/LARS - Purdue Joint Study Program." The second part of the study consisted of development of a closed boundary finding algorithm by IBM and its evaluation by LARS. The purpose of the algorithm is to enable automatic determination of boundaries, such as agricultural

¹LARS (Laboratory for Applications of Remote Sensing) Purdue University

²IBM Scientific Center, Houston, Texas

field boundaries, in digitized aerial or satellite imagery. The algorithm is described and a pictorial evaluation is presented in this report.

II. Field Selection in Digital Images

One of the major problems in the development of automatic pattern recognition systems for earth resources image data is the delineation of closed boundaries around homogeneous areas containing a material of interest. Identification of these closed areas allows categorization of all the enclosed image elements by group pattern recognition techniques. The reliability and speed of the pattern recognition process is markedly increased when groups, i.e., fields, of samples are used. If the fields are not known each image element in the data must be classified separately which is very time consuming and less accurate. Thus, closed boundary finding methods are being widely studied as part of earth resources data analysis research.

Recent work reported by Rosenfield¹ and Anuta² approached the boundary finding problem using the picture gradient which spatially differentiates the data and produces an edge picture. The problem with this approach is that the gradient is inherently "noisy" and produces borders which are discontinuous, of varying width, and the output also consists of many isolated spurious border points. A more stable and lower noise approach using clustering is reported by Wacker.³ The quality of his boundary output is high; however, the algorithm is very time consuming and closed boundaries are not guaranteed. The boundary finding algorithm reported here guarantees closure and is relatively fast. The algorithm description which follows is extracted from an IBM report by Rodd⁴ who developed the program.

III Closed Boundary Finding Algorithm (CBA)

In the following discussion each point of a digital picture is called a pixel (picture element) and a group of points is called a pixel group. In the implementation, pixel groups are squares of g pixels on a side and thus have g^2 pixels in each group. A field is a collection of pixel groups which have been found to come from the same statistical distribution. The algorithm proceeds by computing statistics for a pixel group and comparing them to the statistics of existing sets of pixel groups using the t - test⁵. The t - test is used to determine whether two adjacent groups came from the same distribution and thus from the same field.

1. Statistical Method

Before describing the geometry of building fields using more than one spectral range we describe the statistical method used when comparing two samples. The t -test is used to compare a pixel group which has not been classified as part of a field with some other pixel group or set of pixel groups which form a defined field. The formulas used make two assumptions about the population formed by the pixels: (1) the populations of all fields are normal and (2) the populations of all fields have the same variance.

Call the unclassified pixel group sample 1 and the collection of one or more pixel groups which are classified sample 2. The goal is to compute a value of t to compare against a critical value to determine if the two samples are from a single population (i.e. determine if the pixel group is part of the same field as sample 2).

Let:

x_{ij} be pixel i of sample j ,

\bar{x}_j be the mean of sample j ,

x_{ij} be $(x_{ij} - \bar{x}_j)$,

n_j be the number of pixels in sample j .

Note that $N_j = g^2$ if only one pixel group is considered.

First compute the pooled estimate of variance.

$$s_p^2 = \frac{\sum_i x_{i1}^2 + \sum_i x_{i2}^2}{(n_1-1) + (n_2-1)}$$

Note that the $\sum_i x_{ij}^2$ is computed as usual by

$$\sum_i x_{ij}^2 = \frac{n_j \sum_i x_{ij}^2 - (\sum_i x_{ij})^2}{n_j}$$

Now the value of t is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

A two-tailed test is used with the number of degrees of freedom equal to $(n_1 - 1) + (n_2 - 1)$.

Occasionally a single pixel group will contain pixels from two areas with radically different values of \bar{x}_j or will contain mostly pixels from say a darker area and one or more pixels from an extremely bright area such as a road or a house with a white roof. This will make s_p^2 very large and consequently make t very small thus confirming the null hypothesis that the two samples are

from the same field. Actually this is not the case for sample 1 does not belong with any other pixel groups. In fact, if sample 1 is included in the field with sample 2, any further computations of t using the field with samples 1 and 2 now combined into a single field will produce a large s_p^2 and small t and an acceptance of the null hypothesis. What has really happened is that sample 1 violates assumption 2. To avoid the problem it is necessary to compute the standard deviations σ_1 and σ_2 and if either is greater than some empirical constant times the appropriate \bar{X}_j , the null hypothesis is rejected.

2. Building of Fields

The picture is processed one row of pixel groups at a time. First the field building algorithm will be described in terms of processing a picture of one spectral band. Before processing starts on a row of pixel groups, the sums of the elements and sums of the squares of the elements of each pixel group in the row are computed. The pixel groups are numbered left to right so that s_1 is the sum of elements of the leftmost pixel group in the row and q_1 is the sum of the squares of the elements of the leftmost pixel group. There are W pixel groups in a row.

The k 'th group of the row of pixel groups being processed will be designated as group k . The first group in the current row is thus group 1. The field to which group k is assigned is given by f_k . The k 'th pixel group in the previous row of pixel groups is designated as group (k) . The first such group is group (1) . The field to which group (k) is assigned is given by $(f)_k$.

The first row is easiest to process. Group 1 is arbitrarily assigned to field 1 and field 1 is noted as containing one pixel

group. Letting S_n be equal to the sum of the pixels in field n and Q_n be equal to the sum of squares of pixels in field n , then S_1 is equal to s_1 and Q_1 is equal to q_1 .

Now group 2 is compared to group 1 (group 1 is sample 2 and group 2 is sample 1). If the null hypothesis is accepted, group 2 is added to field 1. Adding to field 1 means noting that field 1 now has 2 pixel groups and

$$S_1 = S_1 + s_2$$

$$Q_1 = Q_1 + q_2.$$

If the null hypothesis is rejected, then group 2 is assigned to field 2 and field 2 is noted as containing one pixel group and

$$S_2 = s_2$$

$$Q_2 = q_2.$$

Next group 3 is compared to group 2 and in general group k is compared to group $k-1$ until the end of the row is reached.

The second row and all succeeding rows are processed by the following steps.

(a) Group 1 is compared to the field containing group (1). Two possibilities exist. Either the null hypothesis is accepted and group 1 is considered part of field $(f)_1$ or the null hypothesis is rejected and group 1 is different.

If they are the same (i.e. group 1 is part of field $(f)_1$), group 1 is added to field $(f)_1$. Adding means noting that field $(f)_1$ has one more pixel group and

$$S(f)_1 = S(f)_1 + s_1$$

$$Q(f)_1 = Q(f)_1 + q_1.$$

Note that $S(f)_1$ means the $(f)_1$ 'th element of S where $(f)_1$ is the field to which group (1), the first group on the previous row, was assigned.

If they are different processing continues to group 2.

(b) Group 2 through $W-1$ are processed in the same manner. Group k is compared to the field which contains group (k) .

If they are the same, group k is added to field $(f)_k$. Then a check is made to see if group $k-1$ has been assigned to a field. If so processing continues to group $k+1$. If not, group $k-1$ is compared to field f_k (which is the same as $(f)_k$). If they are different, processing continues to group $k+1$. If they are the same, group $k-1$ is added to field f_k and then a check is made to see if group $k-2$ has been assigned to a field and so on.

If they (group k and field $(f)_k$) are different a check is made to see if group $k-1$ has been assigned to a field. If not, processing moves to group $k+1$. If so, group k is compared to field f_{k-1} . If they are the same group k is added to field f_{k-1} . If not, processing continues to group $k+1$.

(c) Group W is processed just like groups 2 through $W-1$ except that upon termination of processing of group W , there is no group $k+1$ to begin processing.

(d) After group W is processed, a second pass of the row is made, this one backwards across the row. The pass starts at group W . When a group k is encountered which has not been assigned to a field, group k is assigned to a new field. Then

if group $k-1$ is unassigned, it is compared to group k . If they are the same, group $k-1$ is added to field f_k and group $k-2$ is checked to see if it is unassigned and so forth. Whenever group $k-j$ is unassigned but different from the field which contains groups k through $k-j+1$, then group $k-j$ is assigned to a new field and a check is made to see if group $k-j-1$ is assigned and so forth. Whenever group $k-j$ is already assigned, scanning continues for unassigned groups.

To handle multi-spectral data, a vector S , a vector Q , a vector s and a vector q must be maintained for each spectral band. When two samples are compared, a t is computed for each spectral band. If the null hypothesis is rejected on the basis of the value of t for any spectral band, then the samples are different.

IV. Implementation

The algorithm has been implemented on System/360 under CP-67/CMS. The program is written in FORTRAN with the exception of assembly language subroutines for input data conversion and for timing. The data for each channel are on a separate file on secondary storage. The data files are sequential with each record containing one row of pixels. It is simple to adapt to some other format of raw data because all input of pixel data is done from a single subroutine.

The following parameters are input to the program:

1. The value of g , the number of pixels on one side of a pixel group.
2. Those rows and columns of the picture which are to be processed.

3. The significance level to be used for the t-test.
Four values are available.

4. The number of channels to be processed.

The program can process any number of rows of pixels on a single run. The number of columns is limited by the dimensions of certain variables. Thus, the program is designed to process pictures which are long strips.

Output is to a line printer. Results for the first 50 rows of pixel groups are printed after the first 100 rows are processed and results for the next 50 rows printed after another 50 rows have been processed and so forth. On the output, each field is assigned a character. Only the boundaries of each field are printed. The program is intended to be used with a per field classification program so that when a field is closed (closed means that no pixel groups in the current row of pixel groups belong to that field), the classification program is invoked to classify the closed field as corn etc. Also before printing, any fields in the last row to be printed which are still open need to be classified.

Note that the maximum number of fields which can be open at one time is twice the number of pixel groups in a row. Thus the vectors S and Q need have only $2W$ elements since each time a field is closed, those elements of S and Q which contained the information concerning the field just closed can be released for use by a new field.

A source listing for the program modified to work on the IBM System 360 44PS operating system is included in Appendix I. A discussion of the input data and the program is supplied in Appendix II.

V. Evaluation of the CBA

In order to analyze the performance of the closed boundary finding algorithm the program was applied to two remote sensor data sets. The primary application was to aircraft multispectral scanner data from flights over agricultural areas in Indiana. Regions such as this contain a uniform checkerboard pattern of rectangular fields oriented generally in a north-south direction. Imagery gathered over this area using north-south flight paths will contain a field boundary structure which is colinear with the X-Y axes of the imagery. These conditions make data of this type ideal for testing boundary finding algorithms due to the predictability of the scene border structure. A 1 mile by 8 mile strip of this agricultural scanner data was used to test the CBA using a variety of parameter values for the algorithm.

The CBA was also applied to a second data set to observe the behavior of the algorithm on irregular and randomly shaped features in the imagery. The data set is from a flight over Yellowstone National Park, Wyoming, and contains meadow, forest, rock outcrops and other irregular features.

These test cases are presented as examples to show the typical performance to be expected from the algorithm for these image context categories. The evaluation method is strictly subjective since no reference standard or qualitative evaluation method exists for testing

the boundary results. Field by field visual analysis enables a reasonable judgement to be made as to the quality and consistency of the closed fields the algorithm is finding. Thus, gray scale pictorial printouts and the corresponding closed boundary finding algorithm printouts having the same scale are presented in figure form for documentation and evaluation purposes in this report.

A typical CBA result is presented in Figure 1. A gray scale pictorial computer printout is presented in Figure 1a for a strip of data in the .58 - .65 micrometer band of typical Indiana farm land. This image was generated using a set of 10 print chain characters which reproduce an approximation of a gray scale. The CBA output in Figure 1b was generated using the following parameter values: Minimum points per field = 128; Significance level = 4; Channels = 4,6,8; Pixel group size = 2 by 2. It is difficult to visually evaluate the results on a large scale. Individual fields must be selected in the gray scale printout and the corresponding area located in the CBA output. Inspection of many fields in this manner reveals that the basic border structure of the scene is being captured and in general each field in the gray scale tends to be broken up into two or more subfields. A fixed sequence of symbols is used to represent the closed boundaries and the sequence is repeated as the set of available symbols is exhausted. It is assumed to be desirable to have too many fields defined rather than too few since then there will be less chance of a CBA field covering more than one real field. Also it is entirely reasonable to expect that real fields will contain variabilities which translate to separate sub-fields when observed spectrally.

In order to evaluate the sensitivity of the CBA to variations in the program parameters several figures were generated from CBA output for visual comparison. The minimum number of points allowable for a field is moderately influential in determining the nature of the border output. In Figure 2 three values for MINFLD are illustrated. The difference between MINFLD of 32 and 64 is not great; however, sharper and more linear borders are obtained for MINFLD = 128. The minimum field of 128 picture elements corresponds to 2.4 acres for the scale of this data. Larger minimums may be desirable if the average size of the sub-fields in the scene is larger than this figure. Next, the significance level was varied and Figure 3 contains output for SIGLEV of 2, 3 and 4. It is apparent that this parameter has the strongest effect on the CBA output. For values of 2 and 3 a great deal of border structure is lost and it is assumed that for this scene and three channel operation significance level 4 is the more desirable value to use.

The CBA was originally written to use one channel only to derive borders. The program was expanded to utilize multiple channels and the three channels used (4, 6, 8) tended to give the best results. These channels cover the .52 - .57 (green), .58 - .65 (red) and .72 - .92 (infrared) portions of the spectrum, respectively. These channels tend to be less cross correlated than other combinations of three thus contain more picture information than would channels demonstrating more correlation. Use of more than three channels tended to produce an excess of borders and the output was judged to be degraded. The comparison presented in Figure 4 is for a one-channel versus a three-channel run. It is apparent that a certain amount

of the horizontal border is lost while the vertical border structure appears to be captured relatively well.

These results tended to indicate that it is desirable to use multiple channels for border finding. This approach requires additional computation and the problem of selecting which channels to use is compounded. These considerations led to the decision to test the usefulness of the principal components transformations [6] in the border finding process. This transformation is linear and its effect is to concentrate the variance in the multiple image set in a minimum number of channels. The first principal component is an image channel formed as a linear combination of all original channels such that in this case approximately 75% of the variance in the image set is contained in it. The second principal component contains 12% of the variance, the third 7% and the rest the remaining variance on a decreasing scale. The possible advantages to using the principal components are that fewer channels may suffice and it eliminates the uncertainty of which channels to use.

The principal component transform was performed on the data set which was used to produce the previous output. The CBA was run on the transformed data using the first principal component in one case and the first three in a second case. The results of the runs are displayed in Figure 5 along with the three channel results from Figure 1. Again it is evident that the single channel result tends to miss horizontal boundaries as was observed for the single channel unaltered data. The three component results tend to agree with the

reference case. From this it was concluded that the principal component approach would simplify channel selection but more than one channel would still be required for processing.

The CBA was applied to one other data set obtained from an aircraft scanner flight over Yellowstone National Park, Wyoming, in 1967. A gray scale printout of the .52 - .55 micrometer band of the scanner data is presented in Figure 6. The same parameter values used for previous runs were used for the CBA and the output is pictured on the right in Figure 6. The elongated areas of meadow and rocky material are bordered reasonably well in the center and lower part of the area. The upper area contains forest stands and it is extremely difficult to evaluate what the CBA is doing here. Detailed evaluation of the performance for this scene is beyond the scope of this report and this evidence is presented as an example only.

VI Summary

A closed boundary finding algorithm is described which attempts to extract border information from digitized multispectral imagery. The algorithm was applied to multispectral aircraft scanner data from agricultural and wildland scenes to evaluate its performance. Visual evaluation of results indicate reasonably good performance in the agricultural scene; however, precise evaluation was not possible within the scope of this study. Subjective evaluation of performance for the wildlands case was more difficult than for the agricultural case.

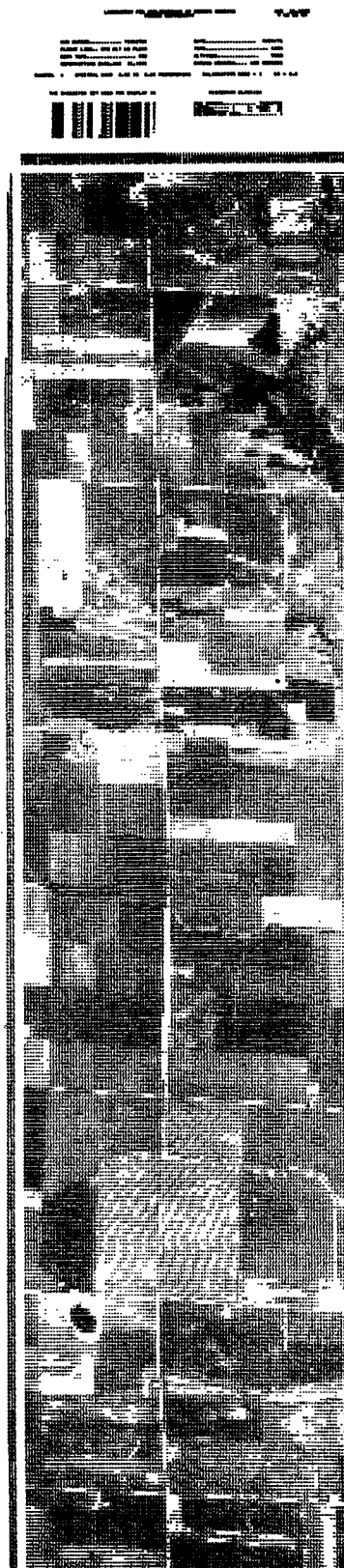
Further work is needed to evaluate and optimize the algorithm. Once satisfactory performance is achieved the output of the CBA must be transferred to a field classification algorithm to achieve the initial intended purpose of the CBA. Classification of the fields could be performed as soon as the fields are closed (completed) or a field output definition could be made and classification performed later. One way this could be accomplished is for the CBA to write a border tape with the fields written in a format easily readable by a per field classifier program.

The processing rate for the present program on an IBM System, 360 Model 67 is approximately .25 sec of CPU time per image line of 222 samples for 1 channel processing and .44 sec per line for three channel processing. The flightlines illustrated averaged 770 lines thus the three channel CBA processing time averaged 339 seconds CPU time. Processing time is not of major concern at this stage since quantitative evaluation of the algorithm is not yet achieved.

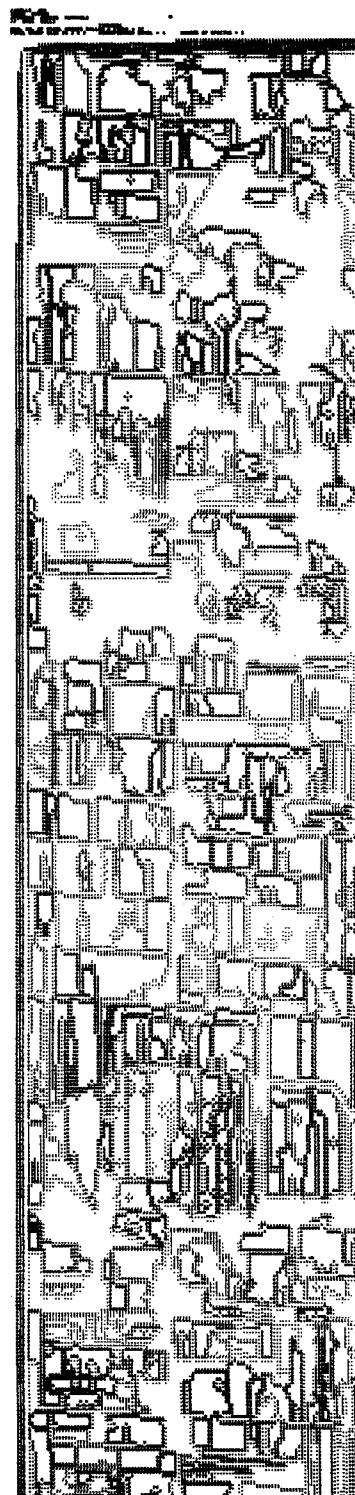
The close cooperation of the IBM Houston Scientific Center and the LARS Purdue organizations was instrumental in achieving the results obtained from this joint study. It is hoped that this document will serve as stimulus to further work on the problems discussed in this report and in the interim report.

References

1. Rosenfeld, A., "Picture Processing by Computer", Academic Press, New York, 1969.
2. Anuta, P. E., "Spatial Registration of Multispectral and Multi-temporal Digital Imagery Using Fast Fourier Transform Techniques", IEEE Transactions on Geoscience Electronics, Vol. GE-8, No. 4, October, 1970.
3. Wacker, A. G., Landgrebe, D. A., "Boundaries in Multispectral Imagery by Clustering", Proceedings of the 9th IEEE Symposium on Adaptive Processes, University of Texas, Austin, December 7 - 9, 1970.
4. Rodd, E. M., "Closed Boundary Field Selection in Multispectral Digital Images", IBM Publication No. 320.2420, IBM Houston Scientific Center, January 14, 1972.
5. Ostle, B., "Statistics in Research", Iowa State University Press, Ames, Iowa, 1963.
6. Ready, P. J. Wintz, P. A., Multispectral Data Compression Through Transform Coding and Block Quantization, "PhD Thesis, School of Electrical Engineering, Purdue University, TR-EE 72-2; and LARS Information Note 050572, May 1972.

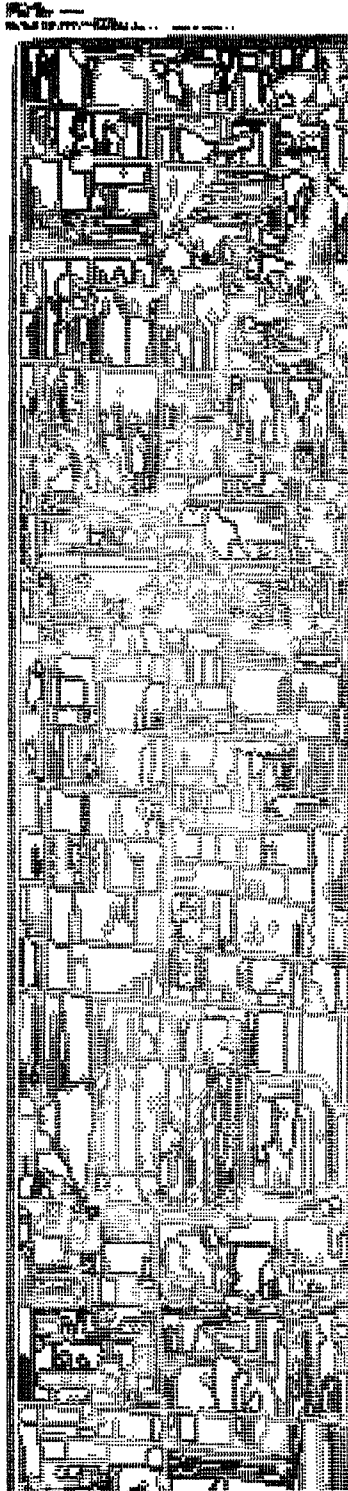


a) Gray Scale Printout

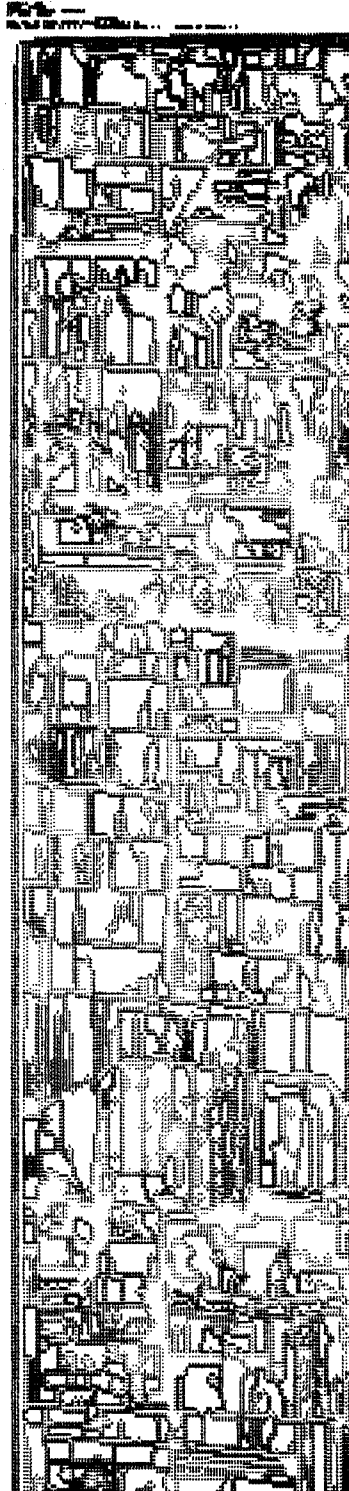


b) CBA Output

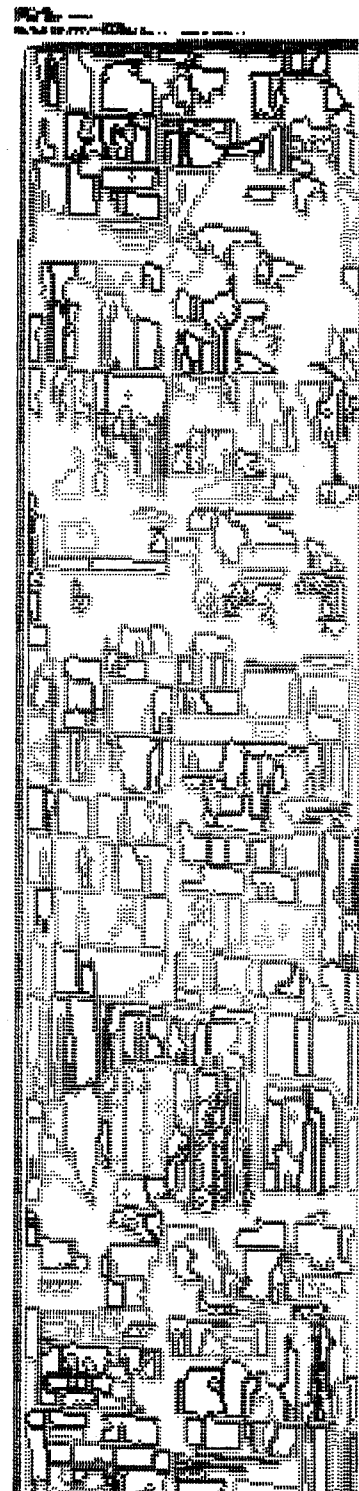
Figure 1. (a) Gray scale line printer image of .58 - .65 micrometer band aircraft scanner data. (b) Closed boundary algorithm output for the scanner data. Parameters: MINFLD = 128, SIGLEV = 4, CHANNELS 4, 6, 8. LARS Run No. 71062701.



MINFLD = 32

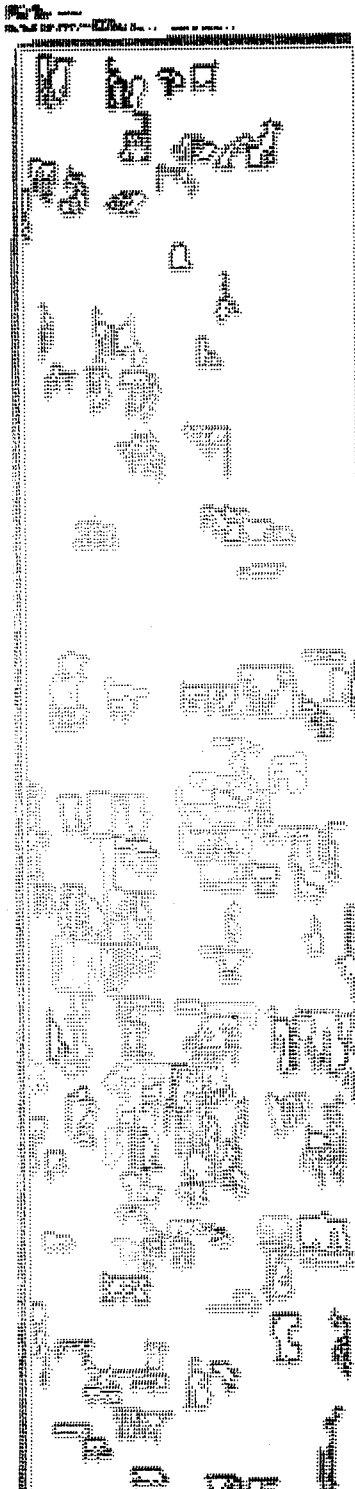


MINFLD = 64

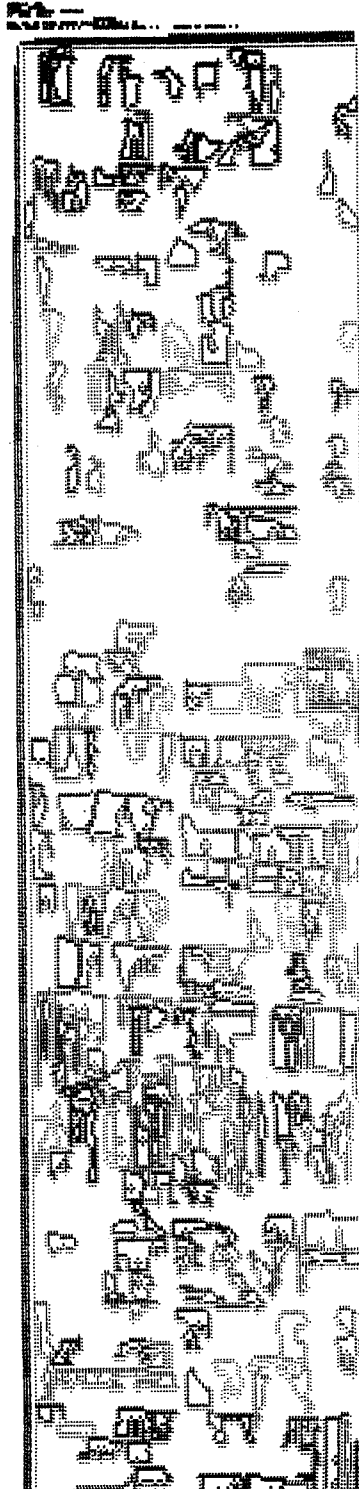


MINFLD = 128

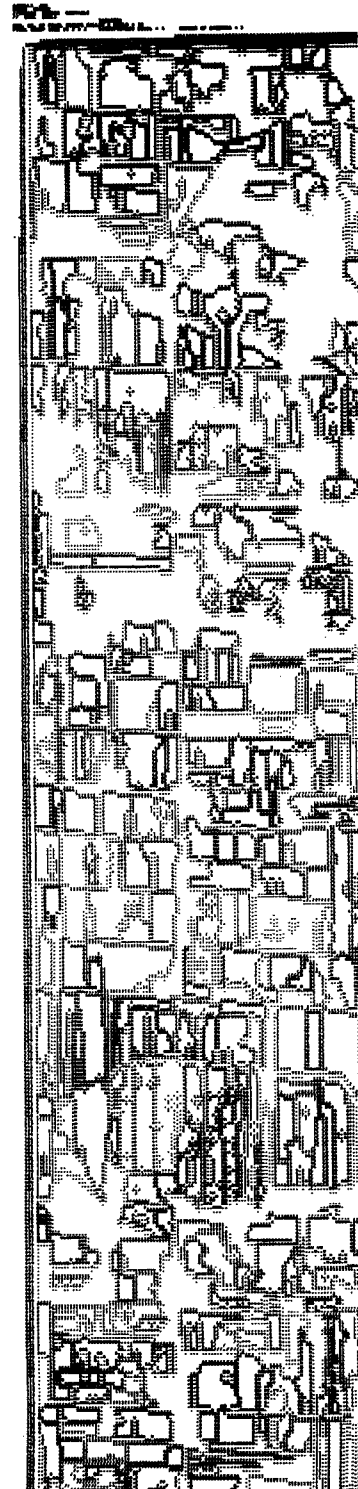
Figure 2. Comparison of the effect of three values of minimum field size on CBA output. Significance Level = 4, Channels 4, 6, 8.



SIGLEV = 2

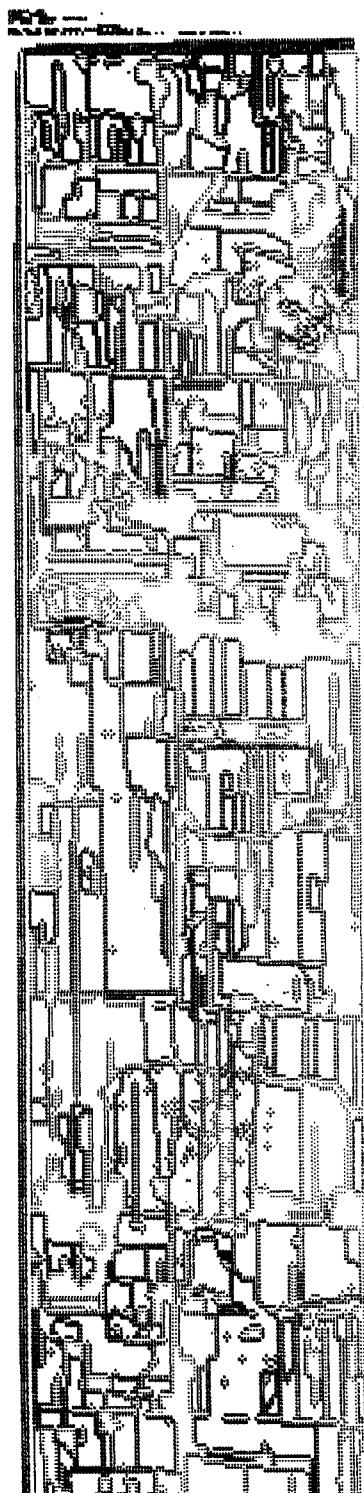


SIGLEV = 3

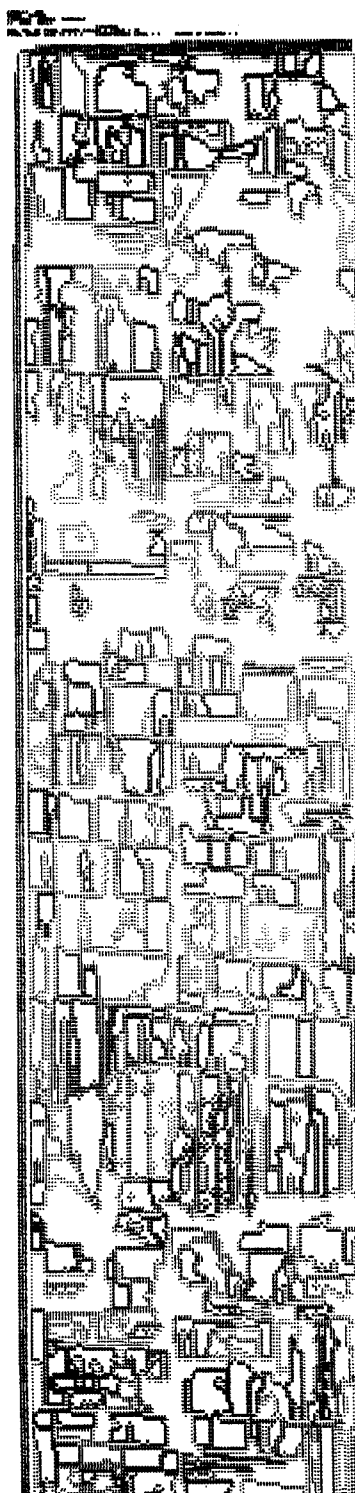


SIGLEV = 4

Figure 3. Comparison of CBA output for three values of significance level. Channels 4, 6, 8 and MINFLD = 128.

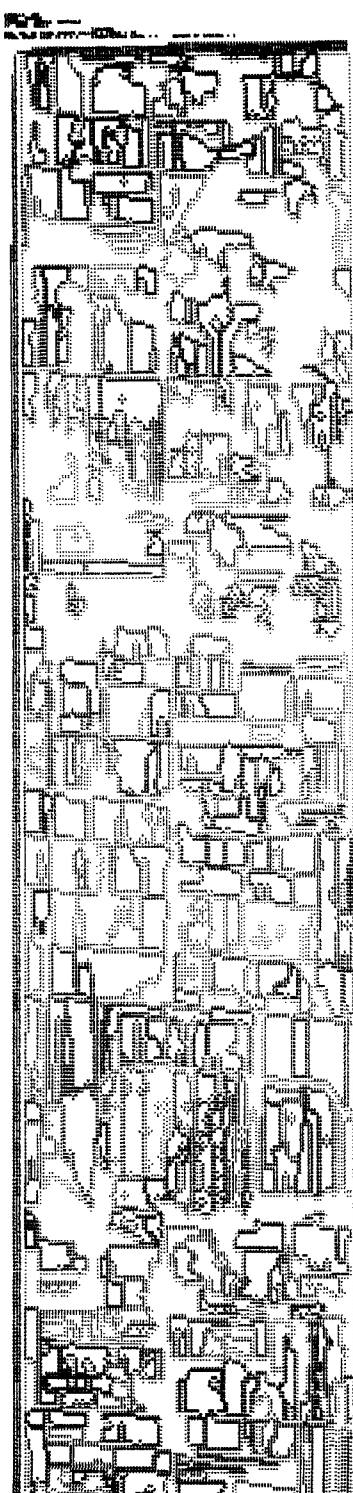


Channel 6

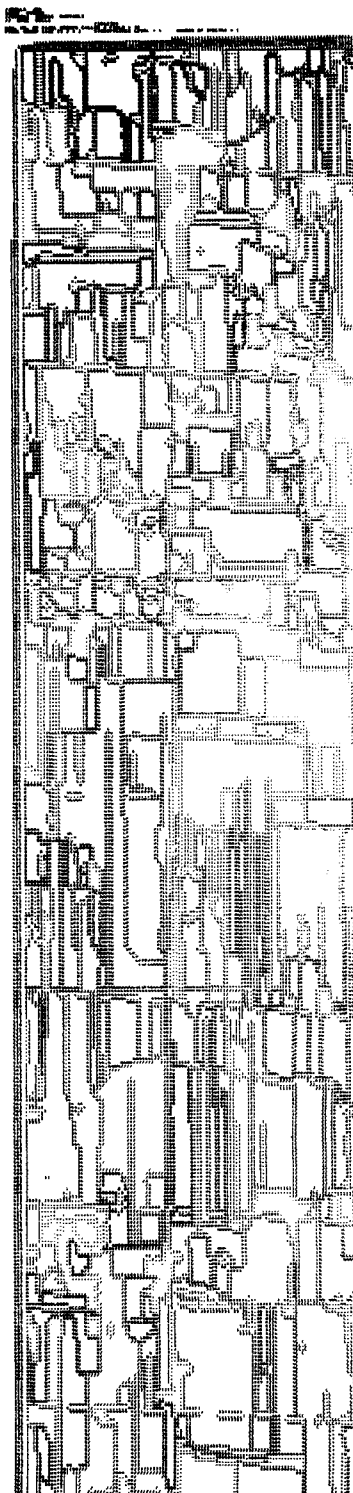


Channels 4,6,8

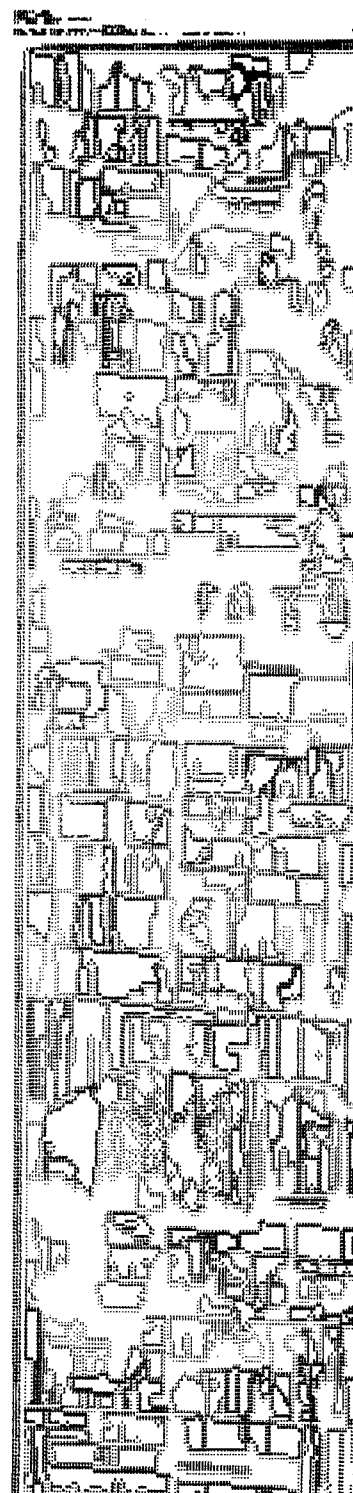
Figure 4. Comparison of CBA output for two sets of channels. Significance level 4 and MINFLD = 128 used.



3 Channels
Untransformed

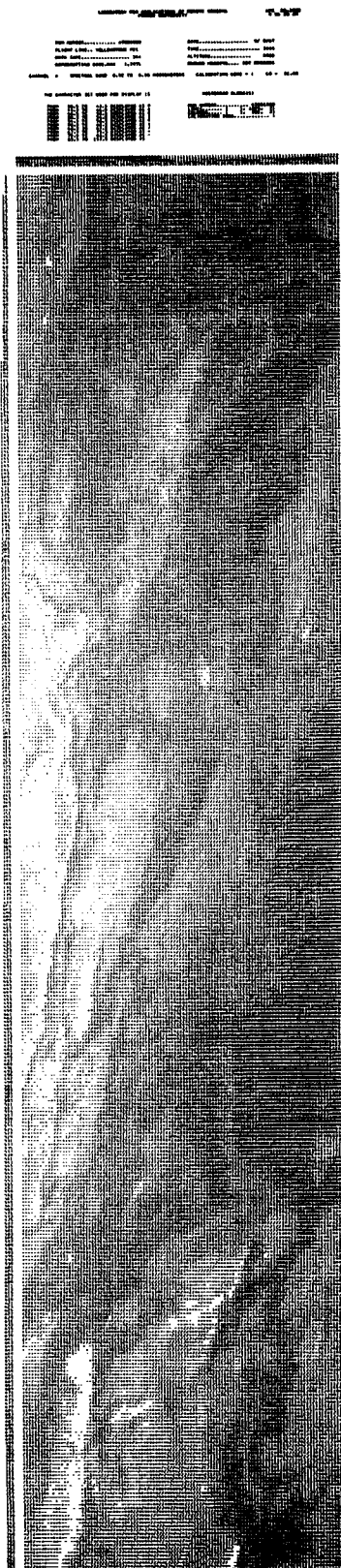


Principal
Component 1

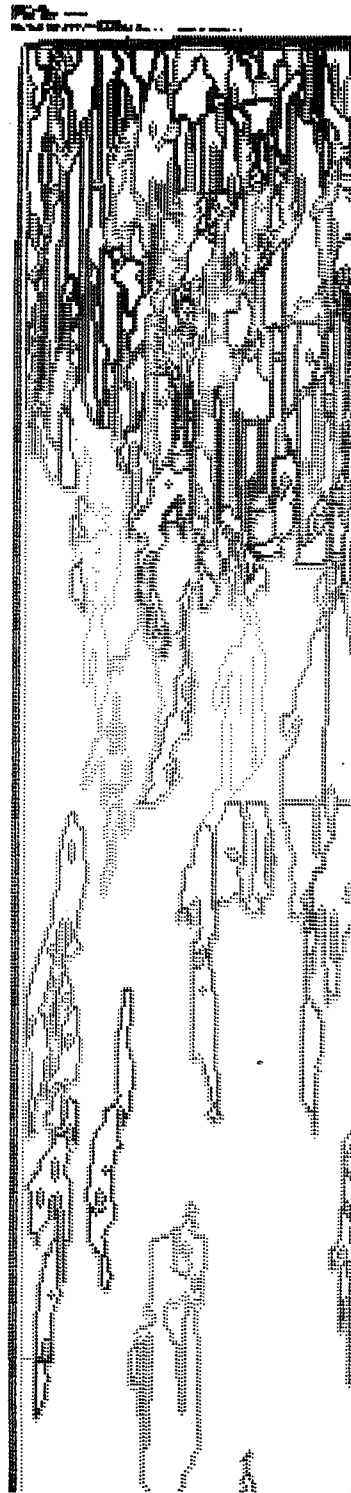


Principal
Components
1,2,3

Figure 5. Comparison of CBA output from unaltered data and principal components data. MINFLD = 128, Significance Level 4.



Gray Scale Printout



CBA Output

Figure 6. Gray scale printout and CBA output for forest and meadow topography in Yellowstone Park, Wyoming. Parameters: MINFLD = 128, Channels 4, 6, 8; SIGLEV = 4.

Appendix I

Fortran Listing of Closed Boundary Finding Algorithm

[illegible]

```

FORTRAN IV      MODEL 44 PS      VERSION 3, LEVEL 4   DATE 7/21/59

0103          IF (ASIN(JW)) GO TO 190                      CEL01570
0104          ASIN(JW) = .TRUE.                             CEL01580
0105          JS = JW                                         CEL01590
0106          FSR(1FLD) = JLINE+1                           CEL01600
0107          RWFLD(JW) = 1FLD                               CEL01610
0108          JW = JW+PGR                                     CEL01620
0109          IF (JW-GE. W1) GO TO 200                       CEL01630
0110          IF (J1 START OF ROW THEN ARE FINISHED.        CEL01640
0111          C                                              CEL01650
0112          IF (ASIN(JW)) GO TO 190                       CEL01660
0113          JS = JW                                         CEL01670
0114          CALL STR                                         CEL01680
0115          C CHECK AGAINST ONE BEHIND IF IT IS NOT ASSIGNED. CEL01690
0116          IF (.NOT. SAME) GO TO 170                     CEL01700
0117          IF A NEW FIELD GO BACK AND ASSIGN.            CEL01710
0118          C                                              CEL01720
0119          ASIN(JW) = .TRUE.                             CEL01730
0120          RWFLD(JW) = 1FLD                               CEL01740
0121          CALL ADD(1FLD,JW)                              CEL01750
0122          C NOW GO CHECK NEXT GROUP BACK ACROSS.        CEL01760
0123          GO TO 180                                       CEL01770
0124          JW = JW+PGR                                     CEL01780
0125          IF (JW-GE. W1) GO TO 170                     CEL01790
0126          C                                              CEL01800
0127          200 CALL PAINT                                  CEL01810
0128          DD 220 I = W1,W2,PGR                          CEL01820
0129          OPENED(RWFLD(I)) = .TRUE.                     CEL01830
0130          CLOSE(RWFLD(I)) = .FALSE.                     CEL01840
0131          IF (.NOT. RWFLD(I)) = .FALSE.                 CEL01850
0132          220 DUE(1) IS MOVED TO CLOSURE FOR COMPARISONS FROM NEXT ROW. CEL01860
0133          ALSO OPEN NOW INDICATE WHICH CLOSURE IS OPENED ON CEL01870
0134          LAST ROW AND CLOSURE INDICATES WHICH VECTORS ARE CLOSED CEL01880
0135          ON CURRENT ROW.                                CEL01890
0136          C                                              CEL01900
0137          DD 230 ICLOSE = 1,250                          CEL01910
0138          IF (OPEN(I)CLOSURE) = .FALSE.                 CEL01920
0139          OPEN(I)CLOSURE = .FALSE.                      CEL01930
0140          230 IF (CLOSURE(I)) = .TRUE.                   CEL01940
0141          FOR ALL FIELDS JUST CLOSED, CLASSIFICATION IS PERFORMED CEL01950
0142          AND THE CLOSURE IS RETURNED TO THE STACK.     CEL01960
0143          THEN OPEN AND CLOSE ARE INITIALIZED FOR THE NEXT ROW. CEL01970
0144          C                                              CEL01980
0145          PRINT IS CALLED TO PRINT RESULTS OF 50 ROWS.   CEL01990
0146          PRINT IS CALLED AFTER EVERY 100 ROWS HAVE BEEN PROCESSED CEL02000
0147          TO PRINT THE FIRST FIFTY ROWS AND THEN AFTER EACH FIFTY ROWS CEL02010
0148          HAVE BEEN PROCESSED TO PRINT AN ADDITIONAL 50 ROWS. CEL02020
0149          IF (JW-GE. W1 .AND. MOD(JNR,50) .EQ. 0) CALL PRINT CEL02030
0150          250 JN = JNR+1                                  CEL02040
0151          C                                              CEL02050
0152          THIS CALL TO PRINT WILL PRINT RESULTS OF ALL REMAINING CEL02060
0153          ROWS.                                           CEL02070
0154          ATIME = FLOAT(1/TIMER(100))/100.0             CEL02080
0155          TIME WILL CONTAIN TIME (VIRTUAL CPU TIME) SINCE THE CEL02090
0156          CALL TO TITER IN STATEMENT 1.                 CEL02100
0157          267 TFORMAT (1,TIME USED=.F8.2*, SECONDS*)    CEL02110
0158          PRINT TFORMAT                                  CEL02120
0159          END                                             CEL02130
0160          END                                             CEL02140

```

```

PORTMAN IV          MODEL 44 PS          VERSION 3, LEVEL 4   DATE 72215

00049  DO 250 JR = IRL,I(R2,PKGR
00050  JNR = JNR +1
00051  C      CALL PGROM
00052  C      JNR IN NEXT ROW OF PIXEL GROUPS.
00053  DO 160 JM = M1,M2,PKGR
00054  ASIN(JM) = .FALSE
00055  C      INITIALIZATION OF PIXEL GROUPS AS ASSIGNED.
00056  JS = JM
00057  C      CHECK AGAINST GROUP ABOVE.
00058  IF (.NOT. SAME) GO TO 130
00059  C      PUT GROUP INTO SAME FIELD AS GROUP ABOVE.
00060  ASIN(JM) = .TRUE
00061  NEW = OLD(JM)
00062  NFIELD = NFIELD
00063  CALL ADDINFLD(JM)
00064  IF (JM -60, M1) GO TO 160

00065  C      CHECK BACK ACROSS FOR UNASSIGNED GROUP.
00066  IF ONE GROUP BACK IS ALREADY ASSIGNED, GO TO NEXT GROUP.
00067  JS = JM-PGR
00068  120  IF (ASIN(JS)) GO TO 160
00069  C      CALL STR
00070  C      COMPARE AGAINST ONE GROUP BACK. IF IT IS DIFFERENT,
00071  C      TRANSFER TO PROCESSING OF NEXT GROUP.
00072  IF (.NOT. SAME) GO TO 160
00073  NEW = OLD(JS)
00074  NFIELD = NFIELD
00075  CALL ADDINFLD(JS)
00076  JS = JS-PGR

00077  C      CHECK TO SEE IF BACK AT BEGINNING OF ROW.
00078  IF (JS -60, M1) GO TO 120
00079  C      IF NO CHECK BACK ACROSS.
00080  130  IF (JM -60, M1) GO TO 160
00081  C      IF NEXT GROUP DIFFERENT FROM THE ONE ABOVE, ASK IF THE ONE
00082  C      TO ITS LEFT ORIGIN IS ASSIGNED AND IF SO COMPARE TO IT.
00083  IF (ASIN(JM-PGR)) GO TO 135
00084  C      IF NO CHECK BACK ACROSS.
00085  IF (JM -60, M1) GO TO 160
00086  C      THE ABOVE IF PROVIDES FOR SPECIAL TREATMENT AT THE END OF A ROW.
00087  GO TO 160
00088  135  C      CALL STR
00089  C      IF (SAME) GO TO 137
00090  IF (SAME) GO TO 137
00091  C      IF THE ABOVE IS AT END OF ROW.
00092  GO TO 160
00093  137  C      ASIN(JM) = .TRUE
00094  NEW = OLD(JM-PGR)
00095  NFIELD = NFIELD
00096  CALL ADDINFLD(JM-PGR)
00097  GO TO 160
00098  140  C      CALL STR
00099  C      SPECIAL CODE FOR NEW FIELD AT END OF ROW.
00100  ASIN(JM) = .TRUE
00101  NEW = OLD(JM)
00102  NFIELD = NFIELD
00103  IF (ONE BACK IS ASSIGNED NO NEED TO COMPARE IT.
00104  JS = JM-PGR
00105  150  IF (.NOT. SAME) GO TO 160
00106  C      DO NOT CHECK BACK ACROSS.
00107  ASIN(JS) = .TRUE
00108  NEW = OLD(JS)
00109  CALL ADDINFLD(JS)
00110  IF (ASIN(JS)) GO TO 160
00111  IF (JM -60, M1) GO TO 150
00112  160  CONTINUE

00113  C      FOLLOWING IS GOING BACK THROUGH ROW TO PICK UP ANY GROUPS
00114  C      NOT ASSIGNED AND CALLING THEN NEW FIELDS.
00115  JM = M1

```

[illegible]

[illegible]

```

FORTRAN IV      MODEL 44 PS      VERSION 3, LEVEL 4   DATE 72215

0001      FUNCTION I(TIMER(J))                                IT100010
0002      ITIMER = 0                                           IT100020
0003      RETURN                                              IT100030
0004      END                                                  IT100040

```

```

FORTRAN IV      MODEL 44  P5      VERSION 3,  LEVEL 4      DATE 72215

0001      SUBROUTINE CLASS
0002      CLASS CLASSIFIES A CLOSED FIELD, THE ELEMENT OF AFCD WITH THE
0003      REVERSE OF THE INDEX AND ALTERS THE PS PRINT OUT
0004      ARRAY TO REFLECT THE CLASSIFICATION
0005      * * * * *
0006      THE DUMMY PER FIELD CLASSIFIER BEGINS WITH THE FIRST EXECUTABLE
0007      STATEMENT AND CONTINUES TO BUT NOT INCLUDING STATEMENT
0008      THE CODE FROM STATEMENT 100 P. 430 ON PROPAGATES THE CLASSIFICATION
0009      TO THE NEXT FIELD OF THE PS PRINTOUT ARRAY WHICH
0010      REPRESENTS THIS FIELD
0011      * * * * *
0012      IMPLICIT INTEGER*2 (E-Z)
0013      INTEGER*4 DUT, DM, N
0014      LOGICAL*4 SAME, ASIN, ENOUGH, CLOSE, OPEN
0015      COMMON /F1F2F3F4F5F6F7F8F9F10F11F12F13F14F15F16F17F18F19F20F21F22F23F24F25F26F27F28F29F30F31F32F33F34F35F36F37F38F39F40F41F42F43F44F45F46F47F48F49F50F51F52F53F54F55F56F57F58F59F60F61F62F63F64F65F66F67F68F69F70F71F72F73F74F75F76F77F78F79F80F81F82F83F84F85F86F87F88F89F90F91F92F93F94F95F96F97F98F99F100F101F102F103F104F105F106F107F108F109F110F111F112F113F114F115F116F117F118F119F120F121F122F123F124F125F126F127F128F129F130F131F132F133F134F135F136F137F138F139F140F141F142F143F144F145F146F147F148F149F150F151F152F153F154F155F156F157F158F159F160F161F162F163F164F165F166F167F168F169F170F171F172F173F174F175F176F177F178F179F180F181F182F183F184F185F186F187F188F189F190F191F192F193F194F195F196F197F198F199F200F201F202F203F204F205F206F207F208F209F210F211F212F213F214F215F216F217F218F219F220F221F222F223F224F225F226F227F228F229F230F231F232F233F234F235F236F237F238F239F240F241F242F243F244F245F246F247F248F249F250F251F252F253F254F255F256F257F258F259F260F261F262F263F264F265F266F267F268F269F270F271F272F273F274F275F276F277F278F279F280F281F282F283F284F285F286F287F288F289F290F291F292F293F294F295F296F297F298F299F300F301F302F303F304F305F306F307F308F309F310F311F312F313F314F315F316F317F318F319F320F321F322F323F324F325F326F327F328F329F330F331F332F333F334F335F336F337F338F339F340F341F342F343F344F345F346F347F348F349F350F351F352F353F354F355F356F357F358F359F360F361F362F363F364F365F366F367F368F369F370F371F372F373F374F375F376F377F378F379F380F381F382F383F384F385F386F387F388F389F390F391F392F393F394F395F396F397F398F399F400F401F402F403F404F405F406F407F408F409F410F411F412F413F414F415F416F417F418F419F420F421F422F423F424F425F426F427F428F429F430F431F432F433F434F435F436F437F438F439F440F441F442F443F444F445F446F447F448F449F450F451F452F453F454F455F456F457F458F459F460F461F462F463F464F465F466F467F468F469F470F471F472F473F474F475F476F477F478F479F480F481F482F483F484F485F486F487F488F489F490F491F492F493F494F495F496F497F498F499F500F501F502F503F504F505F506F507F508F509F510F511F512F513F514F515F516F517F518F519F520F521F522F523F524F525F526F527F528F529F530F531F532F533F534F535F536F537F538F539F540F541F542F543F544F545F546F547F548F549F550F551F552F553F554F555F556F557F558F559F560F561F562F563F564F565F566F567F568F569F570F571F572F573F574F575F576F577F578F579F580F581F582F583F584F585F586F587F588F589F590F591F592F593F594F595F596F597F598F599F600F601F602F603F604F605F606F607F608F609F610F611F612F613F614F615F616F617F618F619F620F621F622F623F624F625F626F627F628F629F630F631F632F633F634F635F636F637F638F639F640F641F642F643F644F645F646F647F648F649F650F651F652F653F654F655F656F657F658F659F660F661F662F663F664F665F666F667F668F669F670F671F672F673F674F675F676F677F678F679F680F681F682F683F684F685F686F687F688F689F690F691F692F693F694F695F696F697F698F699F700F701F702F703F704F705F706F707F708F709F710F711F712F713F714F715F716F717F718F719F720F721F722F723F724F725F726F727F728F729F730F731F732F733F734F735F736F737F738F739F740F741F742F743F744F745F746F747F748F749F750F751F752F753F754F755F756F757F758F759F760F761F762F763F764F765F766F767F768F769F770F771F772F773F774F775F776F777F778F779F780F781F782F783F784F785F786F787F788F789F790F791F792F793F794F795F796F797F798F799F800F801F802F803F804F805F806F807F808F809F810F811F812F813F814F815F816F817F818F819F820F821F822F823F824F825F826F827F828F829F830F831F832F833F834F835F836F837F838F839F840F841F842F843F844F845F846F847F848F849F850F851F852F853F854F855F856F857F858F859F860F861F862F863F864F865F866F867F868F869F870F871F872F873F874F875F876F877F878F879F880F881F882F883F884F885F886F887F888F889F890F891F892F893F894F895F896F897F898F899F900F901F902F903F904F905F906F907F908F909F910F911F912F913F914F915F916F917F918F919F920F921F922F923F924F925F926F927F928F929F930F931F932F933F934F935F936F937F938F939F940F941F942F943F944F945F946F947F948F949F950F951F952F953F954F955F956F957F958F959F960F961F962F963F964F965F966F967F968F969F970F971F972F973F974F975F976F977F978F979F980F981F982F983F98
```

[illegible]

STU00790
STU00800
TU00810
TU00820
TU00830
STU00840
TU00850
TU00860
TU00870
TU00880
TU00890
TU00900
TU00910
TU00920
TU00930
TU00940
TU00950
STU00960
STU00970
STU00980
STU00990

PR100790
PR100800
PR100810
PR100820
PR100830
PR100840
PR100850
PR100860
PR100870
PR100880
PR100890
PR100900
PR100910
PR100920
PR100930
PR100940
PR100950
PR100960
PR100970
PR100980
PR100990
PR101000
PR101010
PR101020
PR101030
PR101040
PR101050
PR101060
PR101070
PR101080
PR101090
PR101100
PR101110
PR101120
PR101130
PR101140
PR101150
PR101160
PR101170
PR101180
PR101190
PR101200
PR101210
PR101220
PR101230
PR101240
PR101250
PR101260
PR101270
PR101280

PRR	0.00	0.10
PRR	0.00	0.20
PRR	0.00	0.30
PRR	0.00	0.40
PRR	0.00	0.50
PRR	0.00	0.60
PRR	0.00	0.70
PRR	0.00	0.80
PRR	0.00	0.90
PRR	0.00	1.00
PRR	0.00	1.10
PRR	0.00	1.20
PRR	0.00	1.30
PRR	0.00	1.40
PRR	0.00	1.50
PRR	0.00	1.60
PRR	0.00	1.70
PRR	0.00	1.80
PRR	0.00	1.90
PRR	0.00	2.00
PRR	0.00	2.10
PRR	0.00	2.20
PRR	0.00	2.30
PRR	0.00	2.40
PRR	0.00	2.50
PRR	0.00	2.60
PRR	0.00	2.70
PRR	0.00	2.80
PRR	0.00	2.90
PRR	0.00	3.00
PRR	0.00	3.10
PRR	0.00	3.20
PRR	0.00	3.30
PRR	0.00	3.40
PRR	0.00	3.50
PRR	0.00	3.60
PRR	0.00	3.70
PRR	0.00	3.80
PRR	0.00	3.90
PRR	0.00	4.00
PRR	0.00	4.10
PRR	0.00	4.20
PRR	0.00	4.30
PRR	0.00	4.40
PRR	0.00	4.50
PRR	0.00	4.60
PRR	0.00	4.70
PRR	0.00	4.80
PRR	0.00	4.90
PRR	0.00	5.00
PRR	0.00	5.10
PRR	0.00	5.20
PRR	0.00	5.30
PRR	0.00	5.40
PRR	0.00	5.50
PRR	0.00	5.60
PRR	0.00	5.70
PRR	0.00	5.80
PRR	0.00	5.90
PRR	0.00	6.00
PRR	0.00	6.10
PRR	0.00	6.20
PRR	0.00	6.30
PRR	0.00	6.40
PRR	0.00	6.50
PRR	0.00	6.60
PRR	0.00	6.70
PRR	0.00	6.80
PRR	0.00	6.90
PRR	0.00	7.00
PRR	0.00	7.10
PRR	0.00	7.20
PRR	0.00	7.30
PRR	0.00	7.40
PRR	0.00	7.50
PRR	0.00	7.60
PRR	0.00	7.70
PRR	0.00	7.80

```

GET00010
GET00020
GET00030
GET00040
GET00050
GET00060
GET00070
GET00080
GET00090
GET00100
GET00110
GET00120
GET00130
GET00140
GET00150
GET00160
GET00170
GET00180
GET00190
GET00200
GET00210
GET00220
GET00230
GET00240

GET00260
GET00270
GET00280
GET00290
GET00300
GET00310
GET00320
GET00330
GET00340
GET00350
GET00360
GET00370
GET00380
GET00390
GET00400
GET00410
GET00420
GET00430
GET00440
GET00450
GET00460
GET00470
GET00480
GET00490
GET00500
GET00510
GET00520
GET00530
GET00540
GET00550
GET00560
GET00570
GET00580
GET00590
GET00600
GET00610
GET00620
GET00630
GET00640
GET00650
GET00660
GET00670
GET00680
GET00690
GET00700
GET00710
GET00720
GET00730

```

FORTRAN IV MODEL 44 PS VERSION 3, LEVEL 4 DATE 72215

```

0062 CALL GADRUN (RUMNUM,1,10,ERRGRN)
0063 IF (ERRGRN.EQ.0) GO TO 335
0064 WRITE(15,2010) ERRGRN
0065 2010 FORMAT(' GADRUN ERROR ',I2)
0066 PAUSE EL BOMBO
0067 335 CONTINUE
0068 DO 340 I = 1,30
0069 340 RESEL(I) = 0
0070 IF (I.EQ.5) WRITE(15,350)
0071 350 FORMAT(' WHICH CHANNELS (412),')
0072 READ (10,360) (PIX(I),I=1,NCHAN)
0073 360 FORMAT(4I2)
0074 DO 370 I = 1,NCHAN
0075 370 RESEL(PIX(I)) = 1
0076 RETURN
0077 END

```

GET00780

GET00790

GET00800

GET00820

GET00840

GET00850

GET00860

FLGS	L-CTR	OBJCODE	ADDR	STMT	SOURCE STATEMENT	
				1 *	SHIFT IS A FORTRAN CALLED ROUTINE TO CONVERT PIXEL DATA TO	XX 00030
				2 *	PROPER INTEGER*2 FORMAT. THE CALLING SEQUENCE IS	XX 00040
				3 *	CALL SHIFT (PIX,DATA,WORDS) WHERE PIX IS AN I*2 VECTOR WHERE	XX 00050
				4 *	THE CONVERTED DATA IS TO BE PLACED. DATA IS A LOGICAL VECTOR	XX 00060
				5 *	WHERE THE DATA COMES IN ONE NUMBER PER BYTE AND WORDS IS	XX 00070
				6 *	THE NUMBER OF NUMBERS TO BE CONVERTED. WORDS IS AN I*2	XX 00080
				7 *	VARIABLE.	XX 00090
000000	00EC	000C	0000C	8	SHIFT CSECT	XX 00100
000000	05C0			9	SR 14,12,12(13)	XX 00110
000004	05C0			10	BALR 12,0	XX 00120
000004	5821	0004	00004	11	USING 0,12	XX 00130
000004	5811	0008	00008	12	L 2,A(1) ADDRESS OF DATA	XX 00140
000004	5811	0008	00008	13	L 2,B(1) ADDRESS OF WORDS	XX 00150
000004	5811	0008	00008	14	L 2,D(1) ADDRESS OF PIX	XX 00160
000012	1855			15 *	SR 5,5 ZERO BXLE COUNTER	XX 00170
000014	18AA			16	SR 10,10 ZERO WDRK REGISTER	XX 00180
000016	5860	C03A	00040	17	L 6,=PI* INCREMENT FOR BXLE	XX 00190
000018	4873	0000	00000	18	LH 7,C(1) LOAD WORDS	XX 00200
00001E	5870	C03A	00040	19	S 1,=PI* GET END OF LOOP CONDITION	XX 00210
000022	43A2	0000	00000	20	LOOP IC 10,0(2) LOAD DATA BYTE	XX 00220
000024	40A1	0000	00000	21	STH 10,0(1) STORE INTO OUTPUT (PIX)	XX 00230
000026	4A20	C03E	00044	22	AH 2,=PI* ADD TO DATA ADDRESS	XX 00240
00002E	4A10	C040	00044	23	AH 1,=M*2 ADD TO STORAGE ADDRESS	XX 00250
000032	8756	C01C	00022	24 *	BXLE 5,0,LDCP	XX 00260
000036	982C	D01C	0001C	25	LH 2,12,28(13)	XX 00270
000038	92F8	000C	0000C	26	BW 14	XX 00280
00003E	07FE			27	BR 14	XX 00290
000040	00000001			28	LTORG	XX 00300
000044	0001			29	DC 2,1	XX 00310
000046	0002			30	DC 4,1	XX 00320
		00000		36	END SHIFT	XX 00340

Appendix II

Input - Output and Program Variable Information

The CBA program listed in Appendix I operates under the IBM System 360 Model 44 programming system 44PS. The first section of this appendix describes the necessary control information, the second describes the format of the picture data used by the program and the third section describes the program output. The fourth section describes the program variables.

I.

All of the control parameters are read from cards by the subroutine GETPRM. Below is a list of (1) the format of the cards, (2) the program variable into which the data is read, and (3) an explanation of the parameter.

1. 'RUN IDENTIFIER' Format is 4A4. AFS

This defines an alphabetic identifier for the run. This identifier will be printed at the head of the output.

2. 'GROUP SIZE' Format is 11. PXGR

This requests the value of g , the number of pixels on the side of a pixel group. Thus a pixel group has g^2 pixels.

3. 'ROWS & COLS 414' Format is 4I4. R1 R2 W1 W2

This requests which rows and columns of the picture are to be processed. Thus it is not necessary to process starting at the top of the picture or at the left of the picture. The data is of the form: first row, last row, first column, last column. If the specified rows and columns are such that they do not come out on a pixel group boundary, the program will delete sufficient rows

and/or columns to come out even. The maximum number of columns which can be processed is 125/(the group size parameter). There is no restriction on the number of rows.

4. 'SIGLEV' Format is I1. SIGLEV

This is the significance level to be used in the t-test. The values of '1', '2', '3', or '4'. A value of '1' will cause the null hypothesis to be rejected most readily and will thus give the least resolution (and cleanest results).

5. 'MIN SIZE' Format is I2. MFS

This requests the number of pixel groups a field must have in order to engage the per field classifier. If you wish to have per field classification performed on all fields, no matter how small, simply punch '01'. Currently this parameter is used only in the dummy per field classifier.

6. 'NUMBER OF CHANNELS' Format is I1. NCHAN

This requests the number of spectral bands to be used in the picture analysis.

7. 'OUTPUT' Format is I1. OUT

This requests the FORTRAN logical data set number to which results are to be written.

8. 'RUN NUMBER' Format I8.

This is the run number of the data set to be processed.

9. 'CHANNELS' Format NI2

Defines the NCHAN channels to be used.

II.

All use of raw digital picture data is in subroutine PGROW. Thus, to use a different format of picture data, only this subroutine must be altered. Subroutine PGROW is called to read in a row of pixel groups. All that PGROW returns are the values of the sums of the pixels in each group (in each channel) and the sums of the squares of the pixels in each group (in each channel). The actual pixel values are not retained. The sum of the elements of the JS'th pixel group in the row for the NS'th channel is placed in ASO(JW,NS). PGROW has available in the named common /INT2/ the number of channel being processed (in NCHAN), the columns of pixels being considered (in W1 and W2) and the size of a side of a pixel group (in PXGR).

III.

The output consists of a heading giving the title, minimum field size, pixel group size, significance level and number of channels processed; a printout of column numbers; and the body of results. In the body of results, the row number is printed at the left. This is the pixel row, not the pixel group row so that if each pixel group is 2 pixels on a side, these numbers on the left will be 1, 3, 5 etc. The rest of the body of results consists of field identification characters. Only the boundaries of each field are printed. Field identification is assigned by a dummy per field classifier in subroutine CLASS. As each field is

closed, this dummy per field classifier gives a new identifier to it. After all the available characters have been used, it starts over at the beginning of the list of characters.

IV.

Meaning of FORTRAN Variables in the Closed Field Selection Program

This section gives the meaning of the important FORTRAN variables in the Closed Field Selection Program. All the variables discussed below except ASP are in a COMMON block. The program uses COMMON to pass virtually all information from subroutine to subroutine.

Control information

OUT contains the FORTRAN logical unit number to which results are written.

IN contains the FORTRAN logical unit number from which control information is read.

NCHAN contains the number of channels being processed.

PXGR is the value g in the paper. PXGR1 is PXGR-1.

R1 and R2 are the first and last rows of the picture to be processed.

W1 and W2 are the first and last columns of the picture to be processed.

MFS is the minimum field size. This is used only in the dummy per field classifier.

Field statistics

Each vector of ASUM is the vector s in the report for one channel. Each vector of ASQ is the vector q in the report for one channel. Each vector of AFTOT is the vector S in the report for one channel. Each vector of AFS in the vector Q is the report for one channel.

AFLD(1) contains the number of pixel groups in the field whose statistics are contained in the 1'th row of AFTOT and AFS.

AT(1) is computed only in the statistical subroutine (subroutine STU) and contains the value of t for the 1'th channel.

ASP is a local variable in subroutine STU and is the value of Sp^{**2} .

Utilization of field statistics variables

STACK is a stack vector. It is initialized so that STACK(1) = 1. NEXT says which element of the stack to be used next. Each time a field is started, a number is pulled from the stack and this number is the number of the element of AFLD and row of AFTOT and AFS to be used for this field. Also when a new field is started, NEXT is increased by one. Each time a field is closed, a number is returned to the stack. Thus, if a field is closed which was described by the 10'th element of AFLD and the 10'th row of AFTOT and AFS, the number 10 is put into the stack and the value of NEXT is increased by one.

Indication of current processing

JR contains the number of row currently being processed. JR is pixel row, not pixel group row. JR is the DO index on the overall loop over rows. If PXGR is 2, then JR is incremented by two each

time through the loop.

JW contains the number of the column currently being processed. JW is the DO index on the loop passing through the row from left to right.

RWFLD(1) contains the element number of AFLD containing information concerning the 1'th pixel group in the current row. RWFLD is actually the vector f in the report.

OLDRW is the RWFLD from the previous row. It is actually the vector (f) in the report.

Printing of results

SP is the array of results. SP is dimensioned (130,100) so that each vector of SP is a line of printed result. After a row is processed, SP is loaded from the vector RWFLD. When a field is closed, and the field is classified, all entries in SP referring to that field are altered to reflect the classification. To tell whether a field has been classified, the program puts the classification index +256 into SP for each element of SP representing the field just classified. After 100 lines have been processed, 50 lines of results are printed and then after another 50 lines are processed lines 51-100 are printed and so forth. Anytime the print routine is called, results are printed from the first 50 vectors of SP. Then the second fifty vectors of SP are moved to become the first 50. As processing continues, further results are placed in vectors 51-100.

LASTR is a vector of dimension 130. It is positioned in COMMON immediately before SP and is thus logically the zero'th vector of SP. This is done to preserve the last row of results just printed.

This information is needed by the print routine to know whether a boundary was crossed.

FSR(1) contains the number of the first vector of SP which represents the field described by the 1'th element of AFLD. FSR is used after per field classification is done to assist in propagating the field classification index (+256) throughout that part of SP representing the field.

JNR contains the total number of rows of pixel groups processed so far. Each time JNR becomes a multiple of 50 (and is at least 100), 50 lines of results are printed.

JLINE is the counter used to index SP.