Reprinted from

Seventh International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Range, Forest and Wetlands Assessment

June 23 - 26, 1981

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

A NEW CLASSIFIER OF MSS DATA - NATURAL BOUNDARY FINDING IN THE FEATURE SPACE

W.Y. CHEN

University of Shandong
Jinan, Province Shandong, China

W.G. COLLINS

University of Aston
Birmingham, England

ABSTRACT

A method of classifying MSS data is proposed which has some advantages in that it does not contain the usual pre-processing constraints, and caters for variation and correlation in the data. In addition it is very economic in terms of computer time.

I. INTRODUCTION

Spectral pattern recognition has been widely used in remotely sensed data processing. Numerous approaches and classifiers have been developed. Among them the following three are typical [1]:

1. Minimum distance to average means classifier,
2. Parallelepiped classifier,
3. Gaussian Maximum likelihood classifier.

These classifications enable a lot of useful work to be undertaken, but they possess some disadvantages. The first one is very clear and simple in concept and it can be done relatively efficiently, but it has an obvious weak point: it does not consider the different variance of different spectral classes. For example, in an unsupervised clustering procedure Euclidean distance between two points in the feature space is often used and all the classes are treated in the same way.

The second classifier has some advantages as it takes account of the fact that different classes have their own particular spectral distribution: the variances are not the same. This classification can be carried out both rapidly and efficiently, but it neglects the fact that the spectral data of ground cover are highly correlated between different bands. If, in the feature space the parallelepiped corresponding to different classes have some overlap, then it would be difficult to discriminate between them.

The third classifier has the advantage in that it takes account of both the spectral data variation and correlation. Currently it is considered as the optimum system - in theory at least - but it takes considerable computer time to use the classification. Moreover, it cannot be claimed that this approach is perfect, even in theory, as the supposition is not always correct that the spectral data of the ground cover is of approximately normal distribution. It is clearly necessary to find another way to carry out data classification which can be done more efficiently and without the drawbacks listed above. The method proposed here is an attempt to do this.

2. THEORETICAL CONSIDERATION AND THE NATURAL CLASSIFICATION PROCEDURE

Using selected MSS bands a feature space can be compiled, each band corresponding to a dimension in it. According to the digital values of a pixel in the relative bands, a representative point can be found in the feature space. If this is done for each pixel in the processed scene, then a set of representative points in the feature space is obtained. Generally speaking, the representative points do not distribute constantly, but at random, rather like the clouds in the sky, somewhere dense and somewhere thin, and it will change from scene to scene. Although the representative points from one particular type of ground cover do not occur in a same position in the feature space, (that is because the variety of the natural conditions) they tend to concentrate together in a small area, to form a cluster. This is the theoretical basis of the three classifiers considered here, and it is also fundamental to the natural classification proposed by this article.

Of course there are some differences. The minimum distance classifier considers that the representative points of a class distribute like a sphere in the feature space, with no correlation between different spectral bands. The parallelepiped classifier takes into account that the representative points of a class distribute in a parallelepiped, which can be of different size corresponding to different ground cover, but it still neglects the correlation between the different bands. The Gaussian Maximum likelihood classifier supposes that the representative points are of multi variate normal distribution.

The method proposed here does not put any constraints on the data distribution, and it is designed to describe the clusters in quite a natural way. The places where there are relatively dense representative points are cluster centres, and the places where there are less representative points are considered as the cluster boundaries. In a two-dimensional feature space if we regard the number of representative points as the height, then the cluster centres are the peaks, and the cluster boundaries are the valleys.

The main procedures are:

1) To incorporate the statistics in the data feature space. (It is possible to include some feature selection or feature transformation, or both, in this step).
2) First to find the class centres, and then find the cluster boundaries by region growing.
3) According to the boundaries to classify the whole scene and to make decisions.

The following are some examples to illustrate how the natural classifier works.

3. CLASSIFICATION USING ONE-DIMENSIONAL DATA

In order to reduce the complexity of the problem and to separate the information from the noise, it is necessary to do some feature selection or feature transformation, or both: if the problem concerned is to find the surface water area, it can be resolved easily by using only one spectral band data. In this case the reflected IR band (band 6 or band 7 in Landsat) is the best choice.

The procedure is as follows:

1) Scan some representative areas to obtain the statistics of a chosen feature.
2) Find all the digital numbers corresponding to the minima of the statistics.
3) Choose the class boundaries from the digital numbers.
4) According to the boundaries to classify the interested area.

Using this method the surface water near the Colchester area (England) can be shown clearly (See Diagram 1).

In some instances the one-dimensional classification can be considered as a special density slice, or a kind of threshold.

4. CLASSIFICATION USING TWO-DIMENSIONAL DATA

The principles are the same as for one-dimensional, but the procedure is slightly different.

1) Scan some representative areas to obtain two-dimensional statistics of the selected two features.

2) Find all the values of the features which correspond to the peaks of the statistics.

3) These peaks will be considered as the cluster centres if they are far enough (the distance relates to the feature space) one from another.

4) Find the cluster boundaries by region growing, and at the same time create an overlay of the two-dimensional statistics. The overlay is using different characters to represent a different class, and can be considered as a two-dimensional class mark table.*

5) Scan the interested area, select the first pixel, refer to the class mark table and classify the pixel. Repeat this process until the whole scene is covered.

The feature selected always corresponds to what is wanted. If it is required to find the vegetation cover of some area then it is not enough to use one spectral band data only. According to the spectral characteristic of vegetation, it may be easier to get satisfactory results by using ratios: e.g. band 4/ band5 and band 7/ band 5.

Diagram 2 gives the statistics in a two-dimensional feature space, and diagram 5 gives the overlay - 'a class mark table'. Diagram 3 is the spectral classification map and 4 is an O/S map of part of Colchester area. The white regions correspond to surface water, and the dark regions correspond to the land; the darker the area, the more likely it is to be vegetation. It is interesting to note that the two bridges across the Abberton Reservoir can be clearly seen.

5. CLASSIFICATION USING THREE-DIMENSIONAL DATA

It is well known that the intrinsic dimensionality of Landsat MSS data is approximately two[2], when dealing with this kind of data, it is not necessary to use more than two dimensions. The following is an example to show how the natural classification works in the multi-dimensional situation.

The principles and the procedures are the same as in the two-dimensional case, but there is a unique feature which must be stressed and which is connected with the computer memory store.

* A Class Mark Table is a table in the feature space which can be considered as an overlay of the corresponding statistics and consists of different characters, each one stands for a definite class. When classifying first find the representative point in the feature space for a pixel, and then use the overlay class mark table to classify that pixel.

It is very difficult to create a three-dimensional feature space as large as 128 x 128 x 128, so besides the feature selection, the feature scale transformation and compression are essential. The simplest example for doing this is shifting and clamping.

The following is a typical procedure:

1) Scan some chosen area to obtain the one or two-dimensional statistics for some features.
2) According to the result of the preceding step make the feature selection, transformation, shifting and clamping in order to compress the data, but keep the necessary accuracy without any degradation.

3) Create the statistics in the three-dimensional feature space.
4) In the space find the points which correspond to the maximum of the statistics.
5) These will, under certain conditions, be accepted as the cluster centre. From these centres the boundaries will be found by region growing, and at the same time an overlay of the statistics will be produced: it is a three-dimensional class mark table.
6) Scan the interested area, select one pixel, look at the mark table; the classification can then be obtained very rapidly. Repeat the procedure for the next pixel, and continue over the whole area.

Diagrams 6 to 8 are examples to show the sections of the three-dimensional statistics and the class mark table, which are through a point in the feature space and parallel to the three coordinates plane.

The spectral classification map produced is similar to 3, but limitations of space and time do not permit further consideration of this aspect.

6. DISCUSSION

The significance of the method proposed here is its efficiency. There are no pre-processing contraints on the data distribution in the feature space, the only supposition is the clustering tendency of the data from one kind of ground cover. When processing, the data of the processed scene is passed through the computer only twice. The first pass is for incorporating the statistics in the feature space: in this pass the only calculation needed is addition, once only for each pixel. The second pass is after cluster boundary finding and is for classification. This pass is very simple and does not need any arithmetic. In this pass the only action is to look at the mark table and obtain the classification result.

Obviously, using this method of classification it is necessary to know about the spectral reflectance of the ground cover, either at the beginning of the processing to extract the feature, or in the final stage to interpret the spectral classification.

7. ACKNOWLEDGEMENTS

All this work was carried out with an ICL 1904S computer, which is not compatible with Landsat CCT's, at the University of Aston in Birmingham, England. Dr. P.D. Mallinson and Mr. R.H. Beswetherick at the computing centre kindly made the tape readable. Dr. P. Carter of Harwell, England gave some useful advice. We would like to express our sincere thanks to them all.

REFERENCES

1. Thomas M. Lillesand, Ralph W. Kiefer, Remote sensing and image interpretation, John Wiley and Sons, New York, 1979.

2. Swain, P.H., and S.M. Davis (eds), Remote sensing: The quantitative approach, Mc.Graw-Hill, New York, 1978.

Diagram 1.

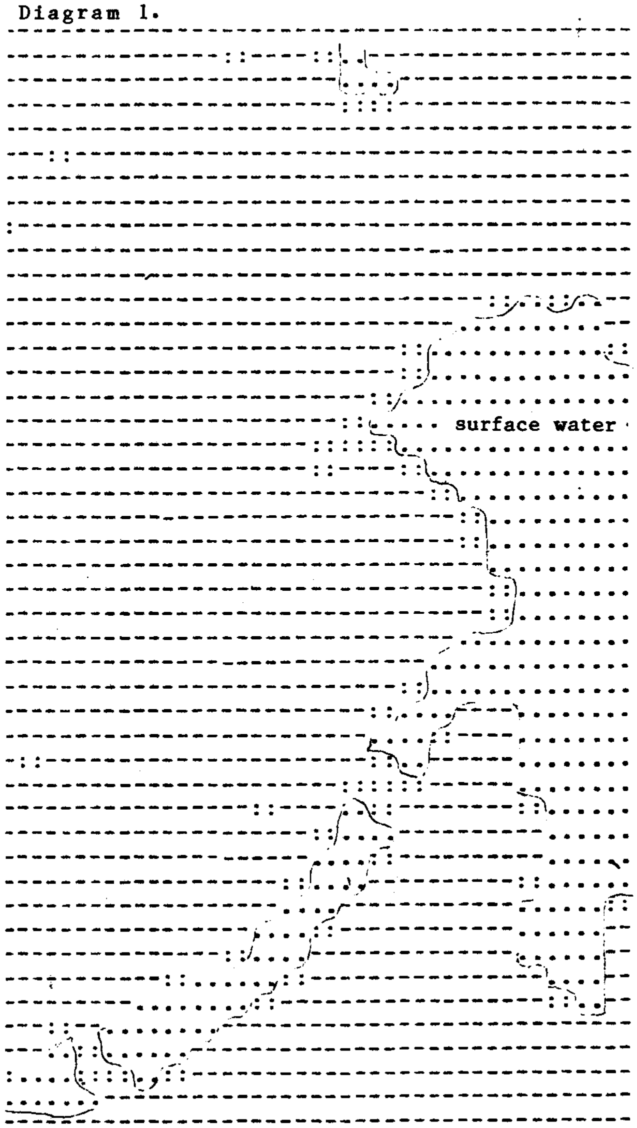


Diagram 2.

THE LOWEST AND HIGHEST I : 23 43
J : 3 50
I & J ARE DIGITAL NUMBER IN TWO MSS BANDS

feature 2 →

```
                      ......::..
                  ..::::::--!!--::..
                  ..::::--TTFFTT!!::..
                ..::----!!FFFFEETT!!--::..
              ....::----!!FFEEEEFFTT!!--::::..
            ..::::--::--!!FFGGGGEEEEFFTT!!--::....
            ..::-----------TTEEWWWWGGGGEEEEFFTT!!--::::..  ..
    ........::--------!!--!!TTFFEEGGFFEEEEEEFFFFFFFFTT!!--  ::........
      ........::::::::........::::::::............
  ........:::::::::::::----!!!!!!TTTT--TTTTFFFFFFFFFFFFTTTT  !!--:::....
  ::::...............::..:::--::--::--::!!----!!!!!!!!  !!----::....
..----:::.....  ........  ......  ........:::..:::----  -----::::::......    ..
::--:::....     ..                              ..    ......  ........
::--::..::..                                              ....  ..:::  ::........
......::::..                                                    ......  ....    ..
      ::....
..  ..  --..
        ..
feature 1
```

WHERE THE MARKS : ,..,::,--,!!,TT,FF,EE,GG,WW, STAND FOR THE PIXEL NUMBER= 5.3*I**2, I=0,1,2,3,....9

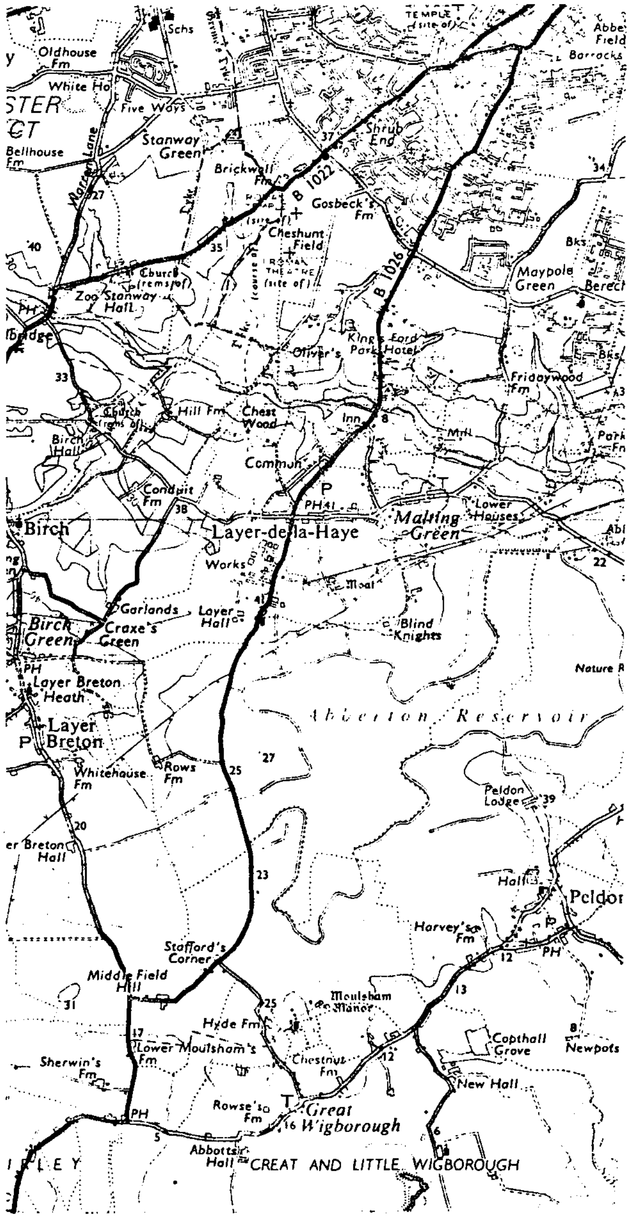
feature 2

0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	3	6	6	11	25	10	2	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	8	26	37	41	79	107	78	25	10	3	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	4	14	59	41	63	190	239	135	112	40	13	3	1	2	0	0	0	0	0	0
0	0	0	0	0	0	0	4	10	23	50	49	94	252	253	264	160	86	59	25	11	5	5	0	0	0	0	0
0	0	0	0	1	0	2	12	13	58	55	60	96	237	283	270	254	160	121	53	33	26	6	3	0	0	0	0
0	0	0	0	1	3	11	23	26	57	45	79	104	230	358	376	327	273	210	172	93	75	44	15	11	2	1	0
0	3	2	2	2	5	12	46	62	66	59	70	68	163	273	428	423	389	362	335	303	220	180	127	73	42	25	13
0	3	8	15	13	20	25	56	70	74	75	106	80	125	162	253	311	359	252	323	303	309	258	239	209	154	92	67
1	5	4	7	10	12	10	32	59	31	25	3	15	12	16	23	30	32	29	13	16	14	11	10	7	1	1	1
1	10	21	19	14	26	26	33	32	45	46	69	71	128	118	131	160	162	79	145	155	251	239	241	240	253	190	155
5	22	45	20	18	13	19	17	14	13	15	25	9	42	41	52	28	49	33	49	33	37	54	30	105	128	129	114
13	49	66	46	27	18	9	4	6	12	3	8	5	12	6	8	5	10	11	11	13	28	22	14	23	43	62	50
34	69	43	32	10	8	3	3	7	1	3	2	0	2	2	0	5	1	2	1	4	2	6	5	2	6	7	6
24	56	30	16	33	17	4	2	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	7	10
7	16	15	43	23	13	4	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	0	2	1	2	1	4
2	5	2	22	12	7	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
6	1	14	2	49	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	3	2	11	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	2	4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
feature 1		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Diagram 3.

```
----FF!!--FF--TTFFFFTTFFFFFFGGWWGGWWGGFFGG
------FFFF!!--TTTTTTTTFF--TTFFFFGGGGGGFFWW
TT------FFFF--TTTTFFTTTTT------FFGGGGFFFFF
------FFGCFFTTFF--TTTT------------GGFFTTTT
------!!FFFF!!TT--FFTT--------TT--GGFFTT--
TT--TT!!FFGGGGFF--TTTT--------TT!!WWGGFFFF
FF!!GGFFFFGGGGFFFFTTTT----GGGGTTGGWWGGGGFF
WWGGGGGGFFGGGGGGFFTTTT--TTWWWWGGFFWWWWGGTT
GGFFFFFF!!--GGWWGGTTTT!!TTFFWWGGGGWWWWWWWW
GGFFFFFFGG--FFGGGG--!!GGTTTTGGGGFFWWWWWWGG
GGFFFFFFFFFFTTFFFF!!FFFFFFFFFFGGGGGGWWWWWW
GGFF--TT----FFFFTTTTGGWWWWFF--!!GGGGWWTTFF
GG!!TT--!!FFGGFFFFTT!!!!FFFFTTTTFFGGFFTTTT
FF--------FFGGFFFFFFTT--!!--TTTTTT!!TTTTTT
FF!!!!----TTTTTTFF!!!!!!FF!!!!!!!!TT--TTFF
!!!!FF!!!!FF--TTFFTTTTTT!!TT--TTTTTTFFFFFF
--TTFF!!!!TTTTTT!!!!FFFFFFTT--FFTTFFFF!!FF
--!!TT----TTTT--TTTTTTFFFFFFFFFFGGFFTTTTTT
----------!!--TT!!TT!!!!FF!!!!FFGGTTFFTTTT
--------------TTTT!!TTTT------------..::!!
--------!!!!!!TT!!TT!!!!!!!!----..
----!!----TTFFTT--FFFF!!------..
----------------FFTTFF--------!!
----------------FFFFFF--------::
!!::----------!!FF--TTFF--!!--::
::  ::--------TTTTTTTT----------
--    ------!!FFFF!!!!TT!!------..
--::::::--TTFFFFTTTT--TT--TT!!TT--
----------FFFFFF--TT--TTTTTTFFTTTT::
----------FFFFFFTTTT!!FFFFTTTTTTTT--
--..------TTFFTTFF!!!!FFTTTTTTTTTT::
--TTTT----------------FFTTTTTTTTTT::
----TT--------!!------!!FFFFTTTTTT::
--------------TT--------------TT--
--------------------!!FFFFTT----
----------::..  ..  !!TTTT!!::
--------::          --FF--::
------::            ::TT--
------..              ::::
----::                            ::::::::
::::                              --!!--!!
::....                          !!--------
::......                    ::!!----!!----
--------..                ::!!!!--!!TT----
----------..              ::!!TTFF--!!--!!
----------..              !!FFFF!!--!!----
--!!--------              ..!!!!!!!!------
------------::                ::!!TT------
------------..                ::!!!!------
--------..                    --TTTT------
------..                      !!!!!!----::
TT..--..                  ::TT!!----------
..--..    ::--            !!!!!!----------
--..    ..----          ::----------------
--!!..  ..!!--        ::------------------
!!....::!!!!!!        --------------------
::    --------::      ..------------------
..    --------..        ::----------------
    ::--!!----::        ::----------------
    !!--------::        ::----------------
  ::!!!!------      ::--------------------
::----------..      ----------------------
!!!!!!--------      ::--------------------
--------------::    ------..--------------
----------------::  ..--------------------
!!----------------..--!!--....--!!--------
```

Diagram 4.



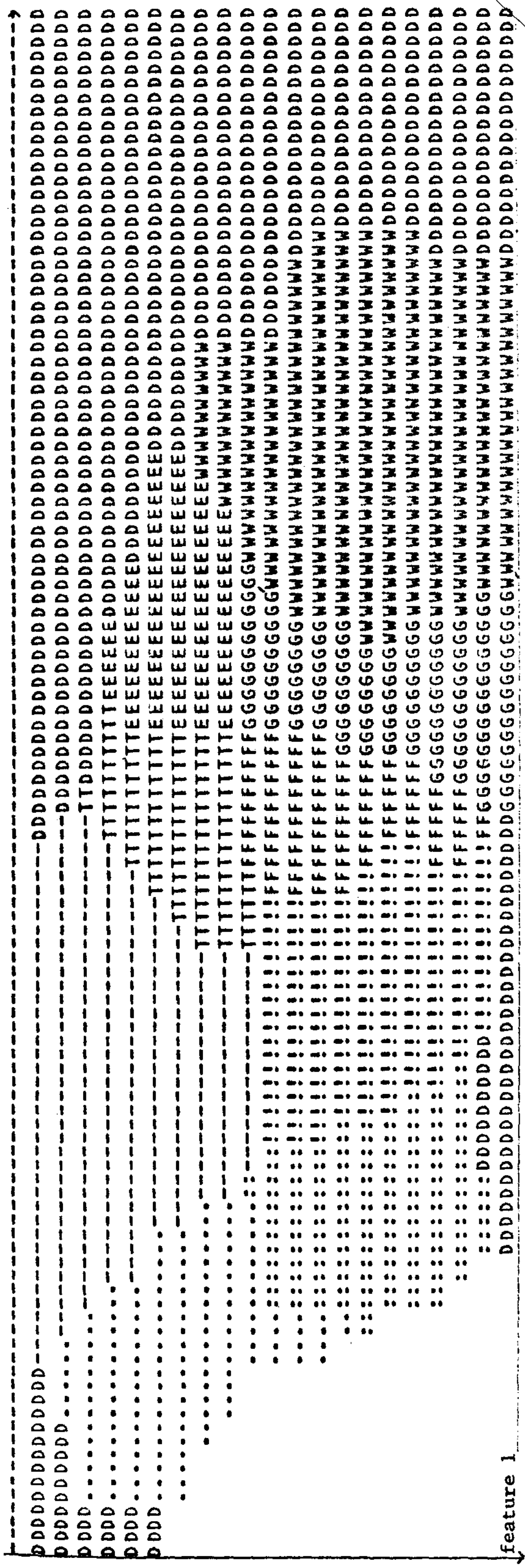


Diagram 5.

Diagram 6.

Digital number in band 4

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	3	2	1	0	0
cluster centre			)	1	·26	17	14	3	0
0	0	0	1	5	27	21	14	5	1
0	0	1	0	6	74	51	17	7	0
0	0	0	13	54	263	193	40	10	0
0	0	0	23	55	366	291	65	9	1
0	0	0	29	110	$\boxed{388}$	338	88	17	0
0	0	3	17	105	301	242	47	5	0
0	0	0	14	85	264	182	51	14	1
0	0	4	13	49	132	94	21	6	0
0	2	10	32	39	73	31	9	2	0
0	4	14	48	51	48	20	6	1	0
0	15	31	50	32	58	11	3	1	0
2	27	62	95	46	46	11	1	0	0
0	40	137	151	77	58	15	4	0	0
0	71	191	195	105	71	18	2	0	0
0	47	168	229	131	54	18	2	0	0
0	24	139	202	116	90	17	3	0	0
0	14	110	186	106	67	14	1	0	0
0	9	55	155	111	60	17	1	0	0
0	2	38	116	91	60	12	1	0	0
0	3	16	102	52	37	11	2	0	0
0	0	11	45	36	19	5	1	0	0
0	0	10	16	23	17	6	0	0	0
0	0	1	10	11	12	2	0	0	0
0	0	1	3	3	6	4	0	0	0
0	0	0	1	8	1	1	0	1	0
0	0	0	0	1	1	0	0	0	0
0	0	1	0	0	0	1	0	0	0

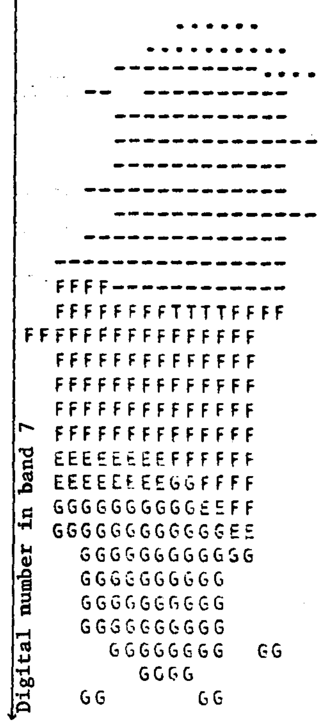


Diagram 7.

Digital number in band 5

63	10	2	0	0	0	0	0	0	0	0
510	163	51	25	17	3	0	0	0	0	0
209	85	87	63	69	26	2	cluster centre			
124	42	51	39	64	27	12	3	0	0	0
42	20	31	53	96	74	16	2	1	0	0
12	20	45	77	200	263	63	11	1	0	0
7	15	32	75	203	366	151	21	7	1	0
7	15	45	90	250	$\boxed{388}$	132	55	10	1	0
2	13	38	77	202	301	164	78	13	3	0
7	13	53	90	213	264	175	86	13	2	1
6	21	40	53	122	132	87	36	17	1	0
8	13	32	31	64	73	44	27	7	2	0
4	8	12	22	49	48	51	32	24	11	3
2	0	9	14	27	58	39	53	40	19	3
2	5	15	10	26	46	40	49	31	41	11
0	6	12	16	41	58	71	33	49	34	23
2	5	12	15	41	71	91	68	57	47	27
1	4	13	11	47	54	98	86	64	45	28
4	6	13	23	52	90	98	79	83	64	49
1	6	15	14	43	67	89	83	69	73	65
1	4	9	16	37	60	77	60	63	56	51
1	2	12	18	36	60	51	36	58	37	35
2	4	8	14	24	37	24	31	35	24	14
1	5	8	10	17	19	14	15	18	11	11
1	4	6	7	18	17	9	12	14	14	7
1	5	10	6	13	12	8	6	7	9	3
2	6	8	6	6	6	6	4	1	2	2
2	2	3	1	1	1	1	1	0	0	0
2	2	3	1	1	1	1	1	0	0	0
2	2	1	2	1	0	0	0	0	0	0

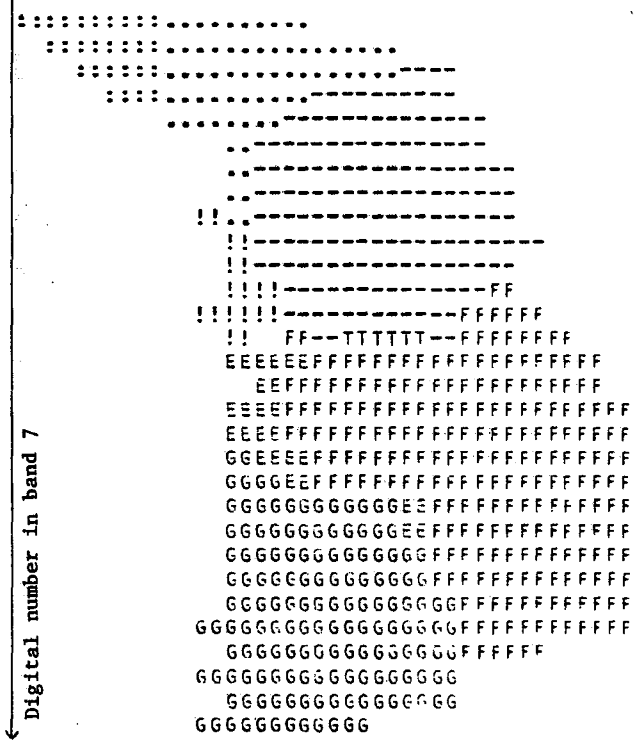


Diagram 8.

Digital number in band 4

0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	3	4	0	0	0	0	0	0	0	0
0	0	1	2	0	0	0	cluster centre			0
0	0	3	1	0	0	0	0	0	0	0
0	1	5	5	1	0	0	0	0	0	0
0	0	2	8	15	8	7	0	0	0	0
0	0	3	2	24	18	15	1	0	0	0
0	0	0	1	14	43	45	10	0	1	0
0	0	0	0	42	105	90	56	2	0	0
0	0	0	4	50	130	250	111	19	2	0
0	0	0	0	29	110	$\boxed{338}$	338	88	17	0
0	0	0	0	7	15	182	179	144	73	4
0	0	0	0	0	3	65	77	122	91	3
0	0	0	0	0	1	10	16	35	51	7
0	0	0	0	0	0	1	1	4	17	1
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0