

# UNSUPERVISED CLASSIFICATION USING

## PIXEL SORT TECHNIQUE

R.H. ALWAN

Scientific Researcher, Remote Sensing Dept., Space & Astronomy  
Research center, Scientific Research Council, Baghdad, P.O.Box  
2441 IRAQ.

F.J.KADUM

Scientific Researcher, Computer Dept., Scientific Documentation  
Center, Scientific Research Council.

G.H. MIZAAL

Assistant Researcher, Remote Sensing Dept., Space & Astronomy  
Research center, Scientific Research Council.

Commission number 3

### ABSTRACT:

---

A new programming technique is proposed for digital image unsupervised classification called the pixel sort technique, depending on sorting of brightness value in the whole image in ascending order, at the same time pixels locations are saved. Then clusters are defined by looking for the similarity of the current pixel using threshold value, starting from the lowest brightness value. By using this sort technique, clusters are easily discriminated and iteration is noticeably reduced. Finally the classified pixels are returned to their original locations. This technique is illustrated by application to a landsat MSS digital image, and a very accurate clustering map is obtained.

## 1. INTRODUCTION

-----

In classification studies it is often desired to know how well the classes can be separated by observing the values of some feature vector for a set of samples. In other words, one wants to know how much information the features provide for distinguishing the classes. To answer these questions, a measure is needed to quantify the amount of information in the feature, SHEN & BADHWAR (1) use Fisher information measure for estimating the mixing proportion of two classes. (Baily & Cowles (2)) apply cluster definition by the optimization of simple measures using complete link method.

When using the statistical approach in performing pattern classification, if the number of features describing the patterns and the number of classes are both rather large, then performance of the usual feature - selection methods and maximum - likelihood (ML) classification can hardly be satisfactory. In such cases the layered classifier (or tree classifier) with the adaptive feature selection function will give an admirable performance but, in general, the design and the optimization of the layered classifier is rather complicated. [WANG RU-YE (3)]. Much work has been done in this field where all ask for a complicated mathematical calculations and consequently this might incur undesirably high demands on computational time. In this paper a new approach is proposed, where the method is based on the sorting technique of digital numbers of a remote sensing digital data. The activities developed in pattern classification have been directed mainly to find identification functions resulting in the classification of data into disjoint classes. This aspect of pattern classification is considered as the process of assigning to each data point a certain degree of belongingness to each class  $C_1, C_2, \dots, C_N$ . away from the complicated algorithms and near to use pure computer science. The general design principle, as well as a concrete description of the method, is given in detail. Classification experiment for remote - sensing image data (LANDSAT MSS data show that the performance of the approach proposed in this paper is good. Classification accuracy is very high and computation time is greatly reduced. In case where both the number of features and the number of classes are large this classifier proved to be very efficient.

## 2. A BRIEF OVERVIEW OF THE SORTING TECHNIQUE:

-----

Sorting technique often used in managing systems which needs certain listing to process, for easier search these listing preferable to be sorted according to code number, alphabetic characters, etc; it depends on the user needs. In scientific problems, sorting technique rarely used.

There are many ways to get sorted data, depends on the type of data. For example, numerique sort, ascending or descending, alphabetical sort, depends on first character only or on the begining two or three characters, etc.

The structure of the sorted file in remote sensing data (see figure 1) is composed of one main pixel and its similar neighbours which have the same value in spectral response or near to it within defined threshold, let the class of the main pixel be  $C_1$ , all pixels belongs to  $C_1$  are garthered in this part of the scene which define one cluster, and the same characteristics are given for  $C_2$ ,  $C_3$  ... $C_N$ .

This method of clustering is an optimization problem which avert the usual methemathical approaches which needs the calculation of the probability density functions of classes as in the (ML) classifier [swain(4)] or claculations of the distance matrix by computing the between - class distance of all the  $k(k-1)/2$  class - pairs among the  $k$  classes as it is in the design of tree classifier [WANG (3)], and others which need long mathematical operations.

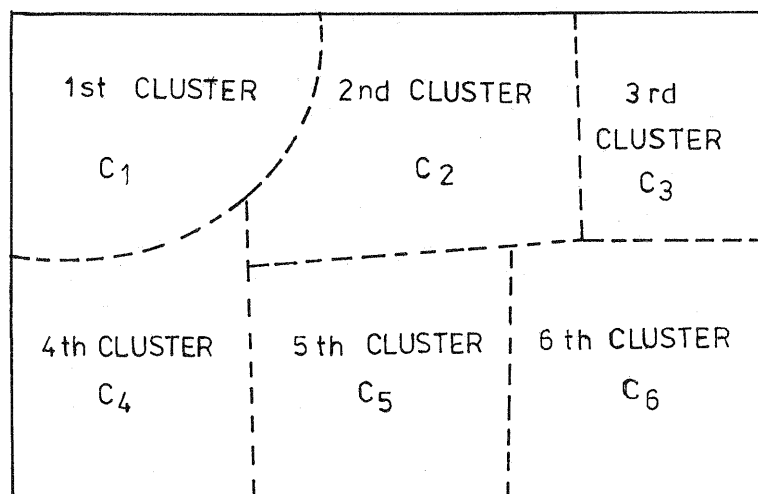


Figure 1. Structure of a sorted file

### 3. CLASSIFIER DESIGN BASED ON SORTING TECHNIQUE:

The classification procedure using the sorting technique is based on rearranging the digital values which represent the spectral response of an image in a form that discrimination between classes and definition of clusters become easy and consequently cluster iteration is greatly reduced.

### 3.1 Reservation of pixels position

The classifier starts by saving the position of each pixel in real number form (I.D) where the integer part (I) indicate the sample number and decimal part (D) indicate the line number, figure 2 shows the new form of the digital image after the addition of pixels positions.

P <sub>11</sub>	1.1	P <sub>12</sub>	1.2	P <sub>13</sub>	1.3	---	P <sub>164</sub>	1.64
P <sub>21</sub>	2.1	P <sub>22</sub>	2.2	P <sub>23</sub>	2.3	---	P <sub>264</sub>	2.64
P <sub>31</sub>	3.1	P <sub>32</sub>	3.2	P <sub>33</sub>	3.2	---	P <sub>364</sub>	3.64
P <sub>641</sub>	64.1	P <sub>642</sub>	64.2	P <sub>643</sub>	64.3	---	P <sub>6464</sub>	64.64

Figure 2. New form of the digital image where each line contains pixel value and its position

This step is done to safe original pixel position before their changes during the sorting procedure.

Each point in an image is considered to belong to one of several mutually exclusive classes [PAIR MAN,(5)], the objective of this work is to put all these classes near to each other. This can be acheived when the sort program is used and applied on the digital data row by row then column by column . To simplify the procedure and reducing computer time, the digital image matrix is transfered to one file contains many records registered sequentially, and each record contains two fields, 1/ pixel value (spectral response), 2/ pixel position . Now the sort program is applied to rearrange the spectral response in ascending order .

### 3.2 Clustering Algorithm

The statistical Pattern - recognition approach considers each measurement to be a realization of a random variable with a fixed class conditional probability distribution defined on the feature space. Ideally, the feature space is chosen so that patterns belonging to the same class and distant from patterns belonging to other classes. This result in the measurment vector from the different classes forming clusters within the feature space. A clustering algorithm attempts to find clusters in a given image and to label each point as belonging to one of the classes thus found. In this paper the classification is done sequentially according to similarity where the main steps of clustering are listed as follows:

1. Every pixel is considered to be a cluster.
2. Define the similarity between the pixel and its neighbours. using certain threshold value (T).
3. Merge the two clusters (first main pixel with similar neighbour pixel to form a new cluster.
4. If neighbour pixel is less than or equal to the main pixel goto (2) otherwise neighbour pixel considered to be a new cluster, then procedure restarts from point (2).

This is repeated until no further changes are made after considering all pixels.

This clustering procedure is demonstrated mathimatically as follows:

Let  $P_c$  be the main pixel considered to be a cluster

if  $V_{cn} \leq V_c + T$  then  $P_{cn} \in P_c$   $n = 1, 2, \dots$  cluster parameter..(1)

where  $V_c$  is spectral response of the main pixel

$V_{cn}$  is spectral response of the neighbours pixel  $P_{cn}$

$T$  is Threshold value

if  $V_{cn} > V_c + T$  then  $P_c = P_{cn}$

From which undefined number of clusters can be obtained and general classificational map or detail classificational map are easily acheived by varying the threshold value (T) in equation (1).

The classes obtained from the sorted file are in the form of clusters arranged sequentially (see figure 1) where the original shape of each cluster have been changed. Returning each cluster to its original shape will necessitate a change in the position of pixels assigned to the cluster.

### 3.3 Return clusters to its orginal shape

It has been shown that the clusters found from above are classified sequentially in other words no similar cluster will be found in other place ; at the same time the produced cluster are not in the orginal shape and position, because they have been changed during the sort procedure. To get each cluster back to its original position, pixels have to return to their original positions, and this will be done by taking the second field which contains the actual pixels position (see figure 2 ) in form of real number where the integer part gives actual row number and decimal part gives actual column number. The classified pixels returs with the class number in which they assign to.

### 3.4 Classification with the sort classifier

with all the work done above, clustering is easy. The very important step is to look for similar pixels which form the cluster, the sort program simplifies the search, because it gathers all the pixels which have the same and the nearby value in one place of the matrix, by this method cluster iteration is greatly reduced. As a result of classification, unknown number of classes is obtained and this number is decreased as the threshold value increased and vice - versa.

### 4. EXPERIMENTAL RESULTS :

-----

The programs which design the mentioned classifier are written in FORTRAN and the classification experiment is performed on an IBM 4341 computer.

This experiment is the classification of the remotely sensed data. Single band of multispectral scanner (MSS) image data size (64 pixels x 64 pixels) subscene have been used. The 4096 pixels are copied to other file with their actual positions. Sort program is applied to sort the spectral response in ascending order, then classification starts by considering the first pixel as the first cluster then procedure continues by looking for the pixels having the same or near spectral response to join it to the cluster. Once the spectral response having higher value than class threshold, the a priori cluster is closed and the current pixel is considered as a new cluster and procedure is repeated as mentioned above. Clustering ends when the last pixel of the data is treated. Finally each classified pixel return to its actual position using the contents of the second field.

The procedure of the sort classifier is shown in figure 3 and the classification output using different threshold value are shown in figure 4.

### 5. CONCLUSION:

-----

Clustering based on sort method leads to good results in discriminating classes, even when the clusters corresponding to the different distribution are close.

The results concerning the classification itself are better than those which need the calculation of statistics parameters like the (ML) classifier and in terms of computer time needed, (ML) classifier was unacceptably long - about 1 hour, while the computer time needed in clustering by sort method was about 10 minutes.

This approach minimizes the probability of error and is based on the idea of sequential clustering and there is no training period before the algorithm starts to classify. Consequently, the algorithm itself sets the classes and the differences between the data; there is no external information which would enable the algorithm to discriminate the data a priori, furthermore, number of classes is not defined a priori. Compared to other approaches (stochastic approximation, dynamic clusters, etc.) this approach is particularly interesting when the data are sequentially observed. Because the data are sequentially treated we have at any instant a partition of the data already observed and the cost of computation is very low.

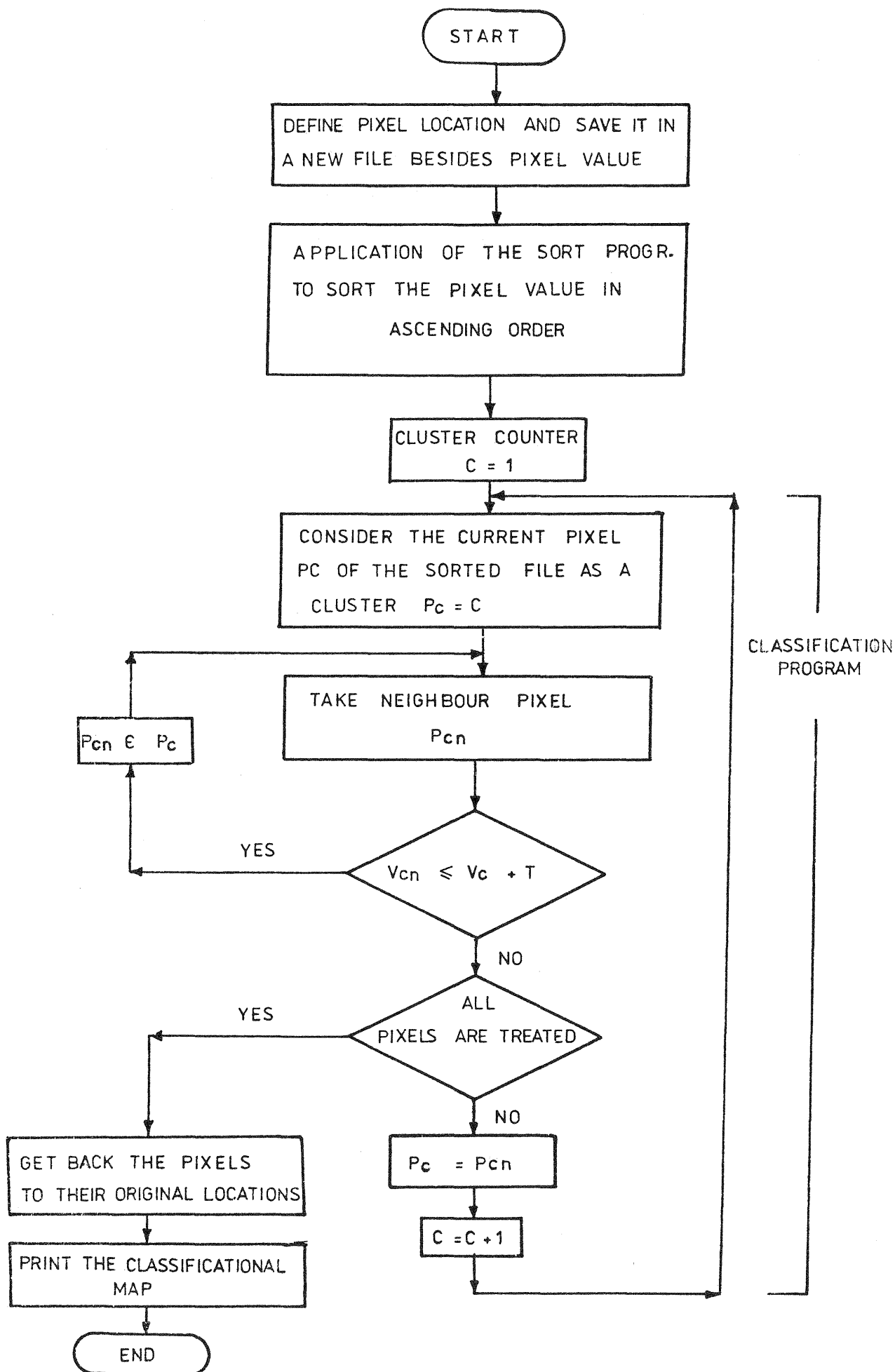
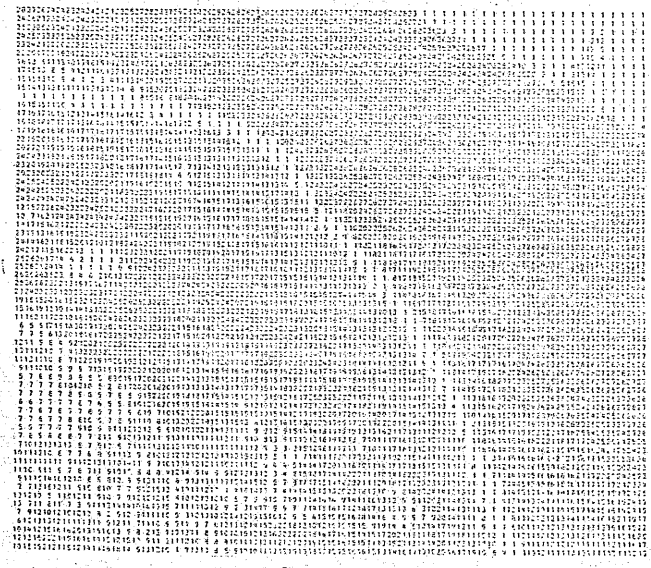


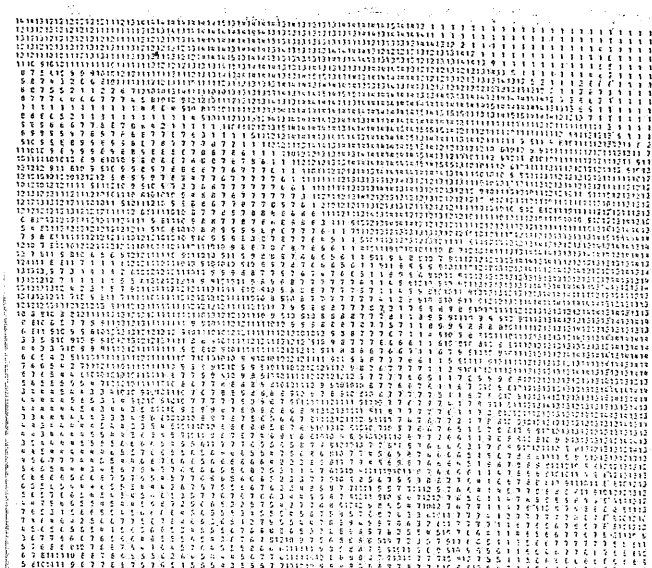
FIG.-3- FLOW CHART OF THE SORT CLASSIFIER



- a -



- b -



- c -

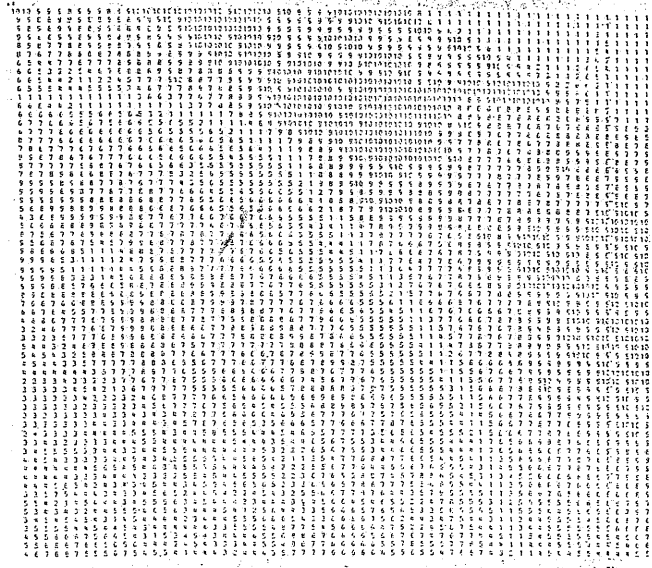


Figure -4- Classification map using sort file technique which gives, a:30 Classes using threshold =5 b:15 Classes using threshold =10 and c:11 Classes using threshold = 15 .

REFERENCES:

- 
1. S.S. SHEN & G.D. BADHWAR, An information measure for class discrimination, INT. J. Remote Sensing, Vol.7 no.4 april 1986 (pp 547 - 556).
  2. THOMAS BAILEY & JOHN COWLES, cluster definition by the optimization of simple measures, IEEE. Trans, on pattern analysis and machine intel., vol. PAMI. 6, no.5 sept. 1984.
  3. WANGE RU - YE, an approach to tree - classifier design based on hierarchical clustering, INT.J. Remote Sensing Vol, 7 no.1. Jan. 1986 pp (75-88).
  4. SWAIN, P.H. & DAVIS S.M. 1978, Remote Sensing, the quantitative approach, New York: McGraw - Hill.
  5. D. PAIRMAN & J.KITTLER, 1986 Clustering algorithms for use with images of clouds, INT. J. Remote Sensing, vol.7 no.7 pp (855 - 866).