

Pattern Recognition with Measurement Space and Spatial Clustering for Multiple Images

R. M. HARALICK, STUDENT MEMBER, IEEE AND G. L. KELLY, MEMBER, IEEE

Abstract—Remote sensor imaging technology makes it possible to obtain multiple images of extensive land areas simultaneously from the radar, infrared, and visible portions of the electromagnetic spectrum. It would be useful to automatically obtain from such data land-use maps indicating those areas of similar types of land, that is, similar as seen through the sensor's eyes.

This classification problem is approached from the perspective of the structure inherent in the data. The classification categories or clusters so constructed are the natural homogeneous groupings within the data. There is high similarity within each cluster and high dissimilarity between clusters.

Two clustering procedures are presented: the first partitions the image sequence and the second partitions the measurement space. In both, the partition is constructed by finding appropriate center sets and then chaining to them all similar enough points. The resulting clusters are simply connected and not necessarily convex.

An example of the measurement space clustering procedure is presented for a set of three multispectral images taken over Phoenix, Ariz.

I. INTRODUCTION

THE pattern recognition problem can be considered as the problem of constructing a classification decision rule and then employing this rule in order to identify a set of measurement N -tuples. In the remote sensing situation, the i th component of each N -tuple can be the measurement made by the i th sensor. For example, in aerial photography a camera with a red filter, a camera with a blue filter, and a camera with a green filter would produce three different components of a measurement N -tuple. Various frequency and polarization combinations can contribute components in a measurement N -tuple for radar imagery.

The set of all possible measurement N -tuples which can be produced from a given set of sensors is usually defined as measurement space. Suppose we denote the N sensors by X_1, X_2, \dots, X_N and the range set for the i th sensor by $L_i, i=1, 2, \dots, N$. Thus sensor X_i produces a measurement which can be any one of the values in L_i .

Formally, measurement space G is defined as the Cartesian product of the range sets: $G = L_1 \times L_2 \times \dots \times L_N$. To facilitate our introductory discussion, we will tentatively use this definition of measurement space.

The classification phase of the pattern recognition problem consists of constructing a decision rule which defines a partition over measurement space, such that each cell of the partition has belonging to it measurements which are similar

to each other. In order to do this, we must have a measure of similarity between any two elements in measurement space.

Similarity is sometimes defined in terms of a priori knowledge. The investigator chooses a set of categories, and he calls similar all those measurements taken of objects in the same category. The problem here is to compromise on the inconsistencies with an appropriate decision rule; such a rule is a Bayes rule [1], [2]. Sometimes similarity is defined in terms of a posteriori knowledge. Two measurements are similar if the data structure indicates that they are similar. In this case the problem is to define a clustering or clumping procedure which links together all those elements which are similar; such a procedure is the multiple linkage clustering method employed in numerical taxonomy [3], [4].

In this paper we define similarity on the basis of the structure of the data themselves. We wish to find the natural groupings or clusters within the data. These natural groupings are sometimes called similarity sets and are characterized by being disjoint subsets of measurement space in which there is high similarity for elements in the same subset and high dissimilarity for elements in different subsets. Since the similarity sets are disjoint and cover measurement space, the collection of similarity sets is a partition over measurement space, each similarity set being a cell of the partition.

Various clustering methods have been developed for constructing such partitions. One strategy analogous to the multiple linkage clustering method employed in numerical taxonomy is as follows: define similarity as being inversely proportional to Euclidean distance and evaluate the similarity between each pair of elements in measurement space. Then in a step-by-step fashion successively link or chain the most similar elements together. However, if measurement space is large, then the time needed to evaluate the similarity for all pairs of elements in measurement space becomes excessive.

A second strategy is an eigenvector technique [5], [6]. Clusters are constructed by determining the dominant eigenvector of the covariance matrix, then projecting the data onto the space spanned by this eigenvector, and splitting them when the projected data have a multimodal distribution. Covariance matrices are found for each of the subsets in the split data and the procedure is repeated until all data subsets have unimodal distributions. The technique is easy to implement, but if the interdispersion of the actual clusters is too large compared to the intradispersion, then it fails.

Manuscript received November 4, 1968; revised January 15, 1969. This work was supported by Department of Defense Advanced Research Projects Agency Work Order 1079 under U. S. Army Engineer Topographic Laboratories Contract DAAK02-68-C-0089. Calculations were carried out by the University of Kansas Computation Center.

The authors are with the Engineering Science Division, The University of Kansas Center for Research, Inc., Lawrence, Kans. 66044.

The ISODATA method iteratively improves the center position of the clusters on the basis of a squared distance criterion [7], [8]. During every iteration, each point is put into the cluster for which the squared distance between it and the cluster center point is least. Then the center point is updated. Although the ISODATA method is much faster than the multiple linkage clustering method, ISODATA is not fast enough for the large amount of points obtained from image data in the remote sensing situation.

Sebestyen has proposed a technique somewhat similar to the ISODATA method [9], [10]. However, the Sebestyen technique is more oriented towards developing a nice representation of the probability density function rather than getting at the cluster structure of the data.

A more extensive survey of clustering techniques may be found in Ball [11] and Friedman and Rubin [12].

The clustering method suggested in this paper is faster than the ISODATA method and is less sensitive to the ratio of cluster interdispersion to cluster intradispersion. It achieves its speed by building one cluster at a time, thus eliminating much of the computation connected with finding the squared distance or similarity of a point with each of the clusters. It is less sensitive to the ratio of interdispersion to intradispersion because examination of the data is made in the full dimensionality of measurement space and not in some one-dimensional subspace. Finally, because clustering is done on the basis of chaining, the method works as easily for long stringy serpentine clusters as for spherical clusters. The clusters formed are simply connected and not necessarily convex.

II. IMAGE DATA

Now let us examine in detail the cluster problem for multiple-image data. Consider first a single image with finite resolution. At first we might conceive of measurement space as the set of all possibly observed 3-tuples (a 3-tuple is an ordered triplet); the first two components are integers specifying the spatial coordinates of the resolution cell and the third component specifies a particular grey density which belongs to the range set of all densities between black and white. However, for any given pair of spatial coordinates there is one and only one measurement in the data set; this is so from any imagery since each resolution cell can have only one grey density. Therefore, we may describe a single image as a two-dimensional sequence of resolution cells, each resolution cell containing a uniform density from the range set as shown in Fig. 1. In this case, it is more useful to conceive of measurement space as the density range set instead of the set of all possibly observed 3-tuples.

Now suppose there are N sensors, each imaging the same environment. The sensors can be cameras, infrared scanners, or radar imagers. To make matters simple we bypass geometric image distortion and congruencing problems and suppose that the images from each sensor are in planimetric correspondence. Let L_i be the range set for image i . L_i is the set of all possible densities which could appear on image i . We may set up the isomorphic correspondence of the lightest density of all the images with the number 0 and the darkest density of all the images with the number 1 and

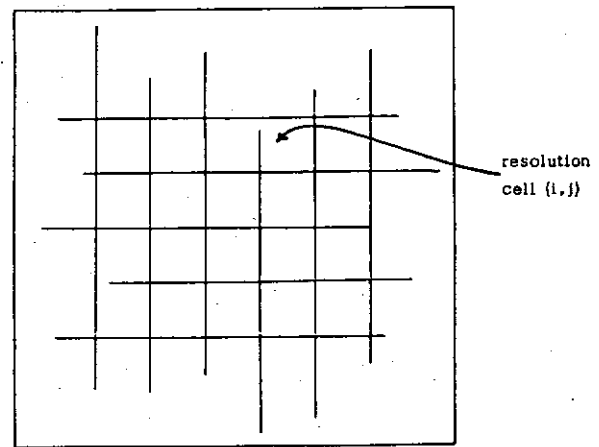


Fig. 1. Single image. Resolution cell (i, j) contains a number which represents the particular grey density which fills the resolution cell.

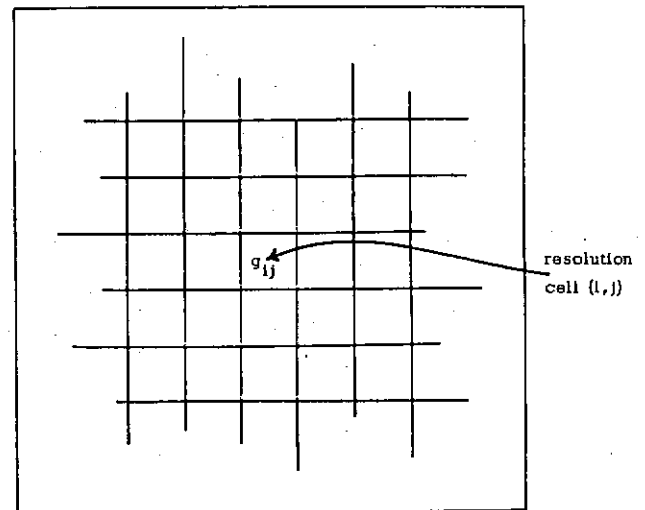


Fig. 2. Multiple image. Resolution cell (i, j) contains the L -tuple density $g_{ij}, g_{ij} = (g_{ij}^1, g_{ij}^2, \dots, g_{ij}^L)$, where g_{ij}^k is the density in the (i, j) resolution cell of the k th image.

all the various intermediate grey densities with the appropriate number between 0 and 1. Thus each range set can be considered as a set of numbers instead of a set of densities. Then each range set is contained in the interval $[0, 1]$.

A multiple image, like a single image, is a two-dimensional sequence where each element in the multiple-image sequence is a N -tuple which is some member of measurement space G . In this case, as mentioned previously, measurement space G is the Cartesian product of the range sets: $G = L_1 \times L_2 \times \dots \times L_N$. If we suppose that the multiple image is rectangular with N_x resolution cells horizontally and M_y resolution cells vertically, then we may represent the multiple image I as the sequence $I = \langle g_{ij} | i \in Z_x, j \in Z_y \rangle$ where $Z_x = \{1, 2, \dots, N_x\}$, $Z_y = \{1, 2, \dots, M_y\}$, and $g_{ij} \in G$ for every i and j , that is, each g_{ij} is an N -tuple whose k th component is some number representing the power received by the k th sensor (we assume power and density to be proportional) when that sensor was looking at the (i, j) resolution cell (see Fig. 2).

Clustering procedures, as we previously mentioned, usually partition measurement space. In the case of image data there are two candidates which might be partitioned.

(5,1)	(5,2)	(5,3)	(5,4)	(5,5)
.20	.25	.18	.67	.65
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)
.15	.17	.22	.72	.64
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)
.16	.14	.17	.69	.75
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)
.32	.34	.31	.31	.29
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
.30	.37	.25	.34	.28

(a) Simple image.

1	2	1	3	3
1	1	1	3	3
1	1	1	3	3
2	2	2	2	2
2	2	2	2	2

(b) Classified image using a best partition of G .

1	1	1	3	3
1	1	1	3	3
1	1	1	3	3
2	2	2	2	2
2	2	2	2	2

(c) Classified image using a best partition of $Z_x \times Z_y$.

Fig. 3.

One candidate is the set G which we have called measurement space, and the other is the set $Z_x \times Z_y$, the set of spatial coordinates. However, a partition of $Z_x \times Z_y$ readily leads to a more natural interpretation of image data as illustrated in Fig. 3(a). Here we have a single image, $Z_x = Z_y = \{1, 2, 3, 4, 5\}$ and $G = \{g | 0 \leq g \leq 1\}$.

Intuitively, a good partition of G based on measurement space closeness with three cells is $H = \{H_i\}_{i=1}^3$, $H_1 = \{g | 0 \leq g < 0.25\}$, $H_2 = \{g | 0.25 \leq g < 0.5\}$, $H_3 = \{g | 0.5 \leq g \leq 1\}$. This partition creates the partition A of $Z_x \times Z_y$ with three cells, $A = \{A_i\}_{i=1}^3$ [see Fig. 3(b)]:

$$A_1 = \{(5, 1), (5, 3), (4, 1), (4, 2), (4, 3), (3, 1), (3, 2), (3, 3)\}$$

$$A_2 = \{(5, 2), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (1, 1), (1, 2), (1, 3), (1, 4), (1, 5)\}$$

$$A_3 = \{(5, 4), (5, 5), (4, 4), (4, 5), (3, 4), (3, 5)\}.$$

A_1 is the set of all spatial coordinates which have measurement space coordinates less than 0.25. A_2 is the set of all spatial coordinates which have measurement space coordinates between 0.25 and 0.5. A_3 is the set of all spatial

coordinates which have measurement space coordinates greater than 0.5.

A good partition of $Z_x \times Z_y$ based on spatial closeness probably is $B = \{B_i\}_{i=1}^3$ [see Fig. 3(c)]:

$$B_1 = \{(5, 1), (5, 2), (5, 3), (4, 1), (4, 2), (4, 3), (3, 1), (3, 2), (3, 3)\}$$

$$B_2 = \{(2, 1), (2, 2), (2, 3), (2, 4), (1, 5), (1, 1), (1, 2), (1, 3), (1, 4), (1, 5)\}$$

$$B_3 = \{(5, 4), (5, 5), (4, 4), (4, 5), (3, 4), (3, 5)\}.$$

This partition of $Z_x \times Z_y$ does not create a partition of G since the resolution cell (5, 2) in B_1 has the density 0.25 and the cell (1, 3) in B_2 also has the density 0.25; hence it cannot be uniquely determined whether 0.25 would go into the cell of a partition of G associated with B_1 or the cell of a partition of G associated with B_2 . We should also note that the resolution cell (5, 2), containing the density 0.25, on the basis of spatial proximity as well as density most likely belongs with group 1, not group 2. It is because of this type of phenomenon that a good partition of $Z_x \times Z_y$ leads to a more natural interpretation than a good partition of G . In this paper we suggest two clustering procedures: 1) spatial clustering, and 2) measurement space clustering. The first method partitions $Z_x \times Z_y$ and the second partitions G .

III. SPATIAL CLUSTERING

Our goal is to describe a clustering procedure which partitions the two-dimensional image sequence I , i.e., partitions the domain of I , $Z_x \times Z_y$. For a given K , where K is the number of clusters, we must find a partition $H = \{H_i\}_{i=1}^K$ such that

$$\bigcup_{i=1}^K H_i = Z_x \times Z_y \quad \text{and} \quad H_i \cap H_j = \emptyset \quad \text{for } i \neq j.$$

From this partition H the classified image C can immediately be constructed. Let $Z_K = \{1, 2, \dots, K\}$. The classified image C is that sequence $\langle c_{ij} | i \in Z_x, j \in Z_y \rangle$ where each $c_{ij} \in Z_K$ and $H_{c_{ij}}$ is the cell of the partition H to which the (i, j) resolution cell belongs. The first step in achieving our goal must be to understand the basis of clustering.

The basis of a clustering procedure is the grouping together of similar items, but what constitutes similarity and why? To answer this question pragmatically we put forth the following model for images of extensive land areas on planetary surfaces:

- 1) things which are very close together are probably the same or similar type of thing
- 2) a sensor which is sensing the same or similar type of object will record the same or similar numerical measurement.

Under this model spatial closeness as measured by Euclidean distance is a good measure of similarity, and we use it as part of the foundation of our clustering method.

To see the rest of the foundation we must examine further the kind of structure we can expect to find in image data. First, relative to all else, the number of resolution cells per

image is very high. This consideration implies that it is out of the question to cluster by comparing the spatial distance and measurement space distance from each resolution cell to every other resolution cell and one by one grouping together those cells closest together. There are too many comparisons. The number of comparisons can be cut down if center sets can be determined. We can cluster by grouping together all those points which are similar enough to the first center set and adding them to the center set. We continue to cluster by grouping together all those points which are similar enough to the now enlarged center set. This procedure can continue until there are no more points similar enough to the first center set. Then we can cluster around the second center set, etc. At this stage all is well; but how do we find the center set? To answer this question we continue our description of the density and spatial distribution structure we can expect to find in image data.

Grossly simplifying, an image may or may not have some sort of homogeneous background(s), and scattered in the background(s) (if any) are various categories of land use or objects. The scattered objects may or may not have a geometric pattern. Each type of object occupies an approximately connected spatial region on the image. The measurement recorded for each point or each resolution cell of each object is not too dissimilar from the measurement recorded at any other point or resolution cell of that same object. In other words, the set of measurements recorded from each object is a fairly homogeneous set. There can be any spatial distribution of objects; one object may only occur once and others hundreds of times or each object may occur approximately the same number of times.

It would be intuitively reasonable to form center sets from those spatial locations which have fairly homogeneous measurement space coordinates and which are representative measurements of a class of objects. However, since the location and extent of objects are unknown to the clustering procedure, it must try to induce this information from the data structure. Since we assume that the set of measurements recorded from any object form a homogeneous set, and the location of these measurements in the image sequence is in a small and more or less spatially connected region, then perhaps by breaking up the image sequence into a set of spatially connected subsequences and examining the measurements in each subsequence we can obtain the necessary information. Thus we make each spatially connected subsequence 1) large enough to include within it a substantial proportion of the measurements recorded from at least one object, and 2) small enough so that a substantial proportion of the measurements recorded from the object make up a large proportion of the measurements in the subsequence. If we can form subsequences in this way, then the empirically observed probability distribution of the measurements in each sequence will be dominated by the substantial proportion of measurements in the sequence recorded from some particular type of object. Thus, if a particular object occurs only once, then there will be one subsequence dominated by it. By picking out the kind of measurements which typify that subsequence (i.e., those

which have high probability in the subsequence), then the set of all the spatial locations containing these measurements is a good center set.

Since the clustering procedure we have proposed starts with center set one, builds on it until not more similar measurements can be found, and then starts building on center set two, etc., we must specify how the order is determined for center sets. We should naturally start with the most important center set and here importance can be correlated with probability. That center set is most important which has the highest probability of all center sets in the subsequence from which it originates.

We now summarize our description. $I = \langle g_{ij} | i \in Z_x, j \in Z_y \rangle$ is the image sequence where $Z_x = \{1, 2, \dots, N_x\}$, $Z_y = \{1, 2, \dots, M_y\}$; $Z_x \times Z_y$ is the domain of the sequence and each g_{ij} belongs to G , the measurement space. I can be represented as a function from the domain $Z_x \times Z_y$ into G ; $I: Z_x \times Z_y \rightarrow G$. Let $S = \{S_n\}_{n=1}^Q$ be a set of spatially connected subsequences of I . By a spatially connected subsequence S_n we mean that for every $g_{ij} \in S_n$ there exists at least one $g_{km} \in S_n$ such that the spatial distance between g_{ij} and g_{km} equals 1, i.e.,

$$d(g_{ij}, g_{km}) = \sqrt{(i-k)^2 + (j-m)^2} = 1$$

and there exists no subsequence S'_n of S_n which has all of its members more than distance 1 away from S_n^c , its complement subsequence in S_n . S_n^c is the subsequence of S_n which contains all the elements of S_n except for those in S'_n .

We define the empirically observed probability $P_i(g)$ on S_i as the proportion of elements in S_i which have measurement space coordinates g . Let $\#$ be the counting measure on $Z_x \times Z_y$ so that, for example, $\#(S_i)$ is the number of elements in the subsequence S_i . The sequence I is a function from $Z_x \times Z_y$ into G ; for each $(i, j) \in Z_x \times Z_y$ there exists one and only one $g_{ij} \in G$ defined by $I(i, j)$, i.e., $g_{ij} = I(i, j)$. Since I is a function, for any $g \in G$, $I^{-1}(g)$ is the inverse image of g ; $I^{-1}(g)$ is the set $\{(k, m)\}$ of all spatial coordinates in $Z_x \times Z_y$ such that $I(k, m) = g$. $\#(S_i \cap I^{-1}(g))$ is then the number of elements in the subsequence S_i which have measurement space coordinates g . P_i is simply defined as

$$P_i(g) = \frac{\#(S_i \cap I^{-1}(g))}{\#(S_i)} \text{ for every } g \in G.$$

From the collection of probability functions $\{P_i(g) | g \in G\}_{i=1}^Q$ we must construct the center sets as well as define their importance. Let W be a function from G into $(0, 1)$, $W: G \rightarrow (0, 1)$, defined by

$$W(g) = \max_j P_j(g) \quad j = 1, 2, \dots, Q$$

and let the sequence $B = \langle g_i | g_i \in G \rangle$ be constructed such that $W(g_i) \geq W(g_j)$ whenever $i \leq j$. The collection of center sets is $\{I^{-1}(g_i)\}_{i=1}^Q$ where $I^{-1}(g_i)$ is more important than $I^{-1}(g_j)$ if $i \leq j$.

There are two parameters which govern how the clustering proceeds. They are K , the maximum number of clusters (i.e., similarity sets or cells in the partition) wanted, and ϵ , a probability cutoff parameter.

We must now describe how the center set is used to form the similarity set. Our description proceeds inductively. For convenience, let $H_0 = \emptyset$. Suppose we have defined H_n . We now define H_{n+1} where n is less than K .

Let

$$t_{n+1} = \min_j \left\{ j | g_j \in G - I \left(\bigcap_{i=0}^n H_i \right) \right\} \quad j = 1, \dots, Q.$$

t_{n+1} is the smallest index of those g_j 's in the sequence B which have not already been included in

$$I \left(\bigcap_{i=0}^n H_i \right).$$

$g_{t_{n+1}}$ is the most important not-yet-used measurement in the set G . Let V_0 be the set of spatial coordinates (i, j) for which $I(i, j) = g_{t_{n+1}}$; that is, it is the set of spatial coordinates on the image which have $g_{t_{n+1}}$ for their measurement space coordinates; $V_0 = I^{-1}(g_{t_{n+1}})$. We construct the similarity set H_{n+1} around its center set V_0 .

First we need the following definitions. Let d be the Euclidean distance metric on $Z_x \times Z_y$:

$$d((i, j), (k, m)) = \sqrt{(i - k)^2 + (j - m)^2}.$$

Let ρ be the Euclidean distance metric on G . For any subset A of $Z_x \times Z_y$, we define the distance between A and the element (i, j) in $Z_x \times Z_y$ as

$$\text{dist}_1((i, j), A) = \min_{(n, m) \in A} d((i, j), (n, m)).$$

Similarly, for any subset D of G we define the distance between D and the element g in G as

$$\text{dist}_2(g, D) = \min_{g' \in D} \rho(g, g').$$

At the first stage we start with V_0 as the center set of H_{n+1} . Then we build on V_0 successively forming $V_1, V_2, \dots, V_{k_0}, \dots$ until there is nothing more which can be added to the current set. We describe how the V_i 's are constructed inductively. Suppose we have defined V_{k_0} ; we now define V_{k_0+1} . Let

$$U = \left\{ (i, j) \in Z_x \times Z_y - \bigcap_{k=0}^n H_k \mid 0 < \text{dist}_1((i, j), V_{k_0}) \leq 1 \right.$$

$$\left. \text{and } 0 < \text{dist}_2(I(i, j), I(V_{k_0})) \leq 1 \right\}.$$

U is the set of all spatial locations (not already included in other similarity sets or in the present set V_{k_0}) which are spatially closer to the set V_{k_0} than 1 and whose corresponding measurement space coordinates are closer in measurement space G to the cluster $I(V_{k_0})$ than 1. If $U \neq \emptyset$, then the chaining procedure is terminated and $H_{n+1} = V_{k_0}$. If not, then let $\{(i, j)\}$ be the set of all spatial coordinates in U , $\{(i, j)\} = U$, such that

- 1) $P(g_{ij}) \geq \epsilon \hat{h}$
- 2) $P(g_{ij}) \leq \hat{h}/\epsilon$

where

$$g_{ij} = I(i, j), \quad \hat{h} = \min_{(k, m) \in V_{k_0}} P(g_{km})$$

and

$$\hat{h} = \max_{(k, m) \in V_{k_0}} P(g_{km}).$$

If $\{(i, j)\}$ is empty, $H_{n+1} = V_{k_0}$. If $\{(i, j)\}$ is not empty, define $V_{k_0+1} = V_{k_0} \cup \{(i, j)\}$. After constructing the disjoint subsets $H_i, i = 1, \dots, K$, we are not sure that

$$\bigcup_{i=1}^K H_i = Z_x \times Z_y.$$

Therefore we initiate a nearest-neighbor spatial search. For every pair of spatial coordinates (i, j) in

$$Z_x \times Z_y - \bigcup_{i=1}^K H_i$$

we assign an index as follows. Let δ be the smallest number such that

$$A = \left\{ (n, m) \mid (n, m) \in \bigcup_{i=1}^K H_i, d((i, j), (n, m)) < \delta \right\}$$

is nonempty. A is the set of all spatial coordinates which already have been classified and which are a distance less than δ away from (i, j) . Let r be the smallest index which maximizes $\#(A \cap H_r)$; we assign the index r to (i, j) . After every pair of spatial coordinates in

$$Z_x \times Z_y - \bigcup_{i=1}^M H_i$$

has been assigned an index r , they are put into the corresponding subset H_r . The union of the H_i 's now covers $Z_x \times Z_y$ and $\{H_i\}_{i=1}^K$ is a partition of $Z_x \times Z_y$.

IV. MEASUREMENT SPACE CLUSTERING

Our goal is to describe a clustering procedure which partitions measurement space G . We proceed in a manner similar to the discussion in Section III. We wish to start with center sets which are ordered. We build on the first center set until no more measurements in G are similar enough to it and then start with a second center set, etc. The difference between the measurement space clustering and the spatial clustering is that in the former, center sets are subsets of $Z_x \times Z_y$ whereas in the latter, center sets are subsets of G .

Let P be the empirically observed probability distribution on G . P is defined by

$$P(g) = \frac{\#(I^{-1}(g))}{\#(I^{-1}(G))}$$

for every g in G , where $I, I: Z_x \times Z_y \rightarrow G$, is the image sequence. The possible center sets will be singleton sets $\{g\}$ for g in G . Perhaps one simple way to order the single sets is by their empirically observed probabilities: the set with the highest probability first and the set with the lowest probability last. However, with such an ordering, some center sets would rank too low—specifically those in small

isolated pockets having relatively high probability compared to their surrounding neighborhoods, but low probability globally. Center sets of relatively high probability locally and low probability globally should rank just below center sets of relatively high probability locally and high probability globally. Therefore, the ordering for the elements in measurement space G , obtained from the empirically observed probability function P , must clearly be based on more global properties than the probability of $\{g\}$.

We propose that the ordering be based on the association of the cylinder sets which characterize the singleton sets $\{g\}$ for g in G . Suppose the range sets L_i are quantized so that $L_i = \{l_{i1}, l_{i2}, \dots, l_{iN_i}\}$, $i = 1, 2, \dots, N$. The cylinder set E_{ij} is defined by

$$E_{ij} = \{g = (x_1, x_2, \dots, x_N) | x_i = l_{ij}\};$$

$$i = 1, \dots, N; j = 1, \dots, N_i.$$

For every $g \in G$, $g = (x_1, x_2, \dots, x_N)$, let $\delta(g)$ be the characteristic function of g ; $\delta(g) = (\delta_1(g), \delta_2(g), \dots, \delta_N(g))$, where $\delta_i(g) = j$ if and only if $x_i = l_{ij}$. It can easily be shown that for any $g \in G$,

$$\{g\} = \bigcap_{i=1}^N E_{i\delta_i(g)}$$

Since each singleton subset $\{g\}$ is equivalent to an appropriate intersection of cylinder sets, and the cylinder sets do reflect more global properties than $\{g\}$, it is natural to base the ordering on the association between the appropriate cylinder sets.

We use the coefficient V as a measure of association [13]. For any two subsets A and B of G , V is defined as follows:

$$V(A, B) = \frac{P(A \cap B)P(A^c \cap B^c) - P(A^c \cap B)P(A \cap B^c)}{[P(A)P(A^c)P(B)P(B^c)]^{1/2}}$$

where A^c is the complement of A . It is quickly verified that V has the following properties:

- 1) $V(A, A) = 1$
- 2) $V(A, B) = V(B, A)$
- 3) $V(A, B) = 0$ if and only if $P(A \cap B) = P(A)P(B)$
- 4) $V(A, B^c) = -V(A, B)$.

Let f be the function which establishes the rank of each g in the ordering:

$$f(g) = \max_i \quad i = 1, \dots, N \quad V\left(E_{i\delta_i(g)}, \bigcap_{\substack{j=1 \\ j \neq i}}^N E_{j\delta_j(g)}\right).$$

The sequence $B = \langle g_i | g_i \in G \rangle$ is constructed such that $f(g_i) \geq f(g_j)$ whenever $i \leq j$.

As in Section III, two parameters govern how the clustering proceeds: K , the maximum number of clusters (i.e., similarity sets or cells in the partition) wanted, and ϵ , a probability cutoff parameter.

We now may describe the clustering procedure inductively. For convenience, we let the first cell or cluster H_0 be the empty set ϕ . Suppose we have defined cell H_n . We define H_{n+1} for n less than K .

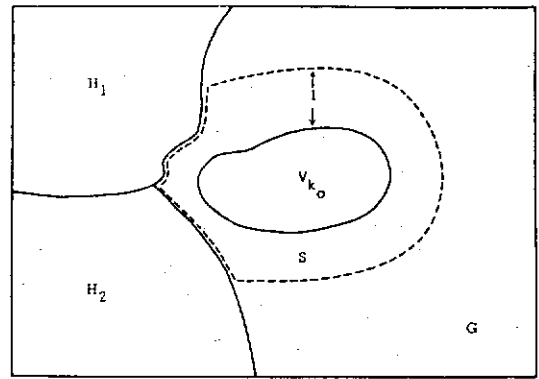


Fig. 4. Geometric illustration of how distance is used in the clustering procedure.

Let

$$t_{n+1} = \min_j \left\{ j | g_j \in G - \bigcup_{i=0}^n H_i \right\}.$$

Thus t_{n+1} is the smallest index of those g_j 's in the sequence B which have not already been included in

$$\bigcup_{i=0}^n H_i,$$

and $\{g_{t_{n+1}}\}$ is the most important unused center set contained in G . We construct the cluster H_{n+1} around its center set $V_0 = \{g_{t_{n+1}}\}$. We build on V_0 , successively, forming $V_1, V_2, \dots, V_{k_0}, \dots$, until there is nothing more which can be added to the current set. We describe how the V_i 's are constructed inductively. Suppose V_{k_0} has been defined. We assume that there is a metric ρ on G ; $\rho: G \times G \rightarrow [0, \infty]$. Therefore, ρ [14] has the following properties:

- 1) for every $g_1, g_2 \in G$, $\rho(g_1, g_2) \geq 0$ with equality if and only if $g_1 = g_2$
- 2) for every $g_1, g_2 \in G$, $\rho(g_1, g_2) = \rho(g_2, g_1)$
- 3) for every $g_1, g_2, g_3 \in G$, $\rho(g_1, g_3) \leq \rho(g_1, g_2) + \rho(g_2, g_3)$.

In this paper we take ρ to be the standard Euclidean distance measure and assume that each of the values in the range sets are integers. Let

$$S = \left\{ g \in G - \bigcup_{i=0}^n H_i | 0 < \text{dist}(g, V_{k_0}) \leq 1 \right\}$$

where

$$\text{dist}(g, V_{k_0}) = \min_{g' \in V_{k_0}} \rho(g, g').$$

S is the set of all the elements in G which are not already included in other clusters

$$\bigcup_{i=1}^n H_i$$

or the present set V_{k_0} , and which are a distance less than 1 away from V_{k_0} , as illustrated in Fig. 4. If $S = \phi$, then clustering procedure is terminated for the $(n+1)$ th similarity set, and we define $H_{n+1} = V_{k_0}$. If we find a $g \in S$ which is similar

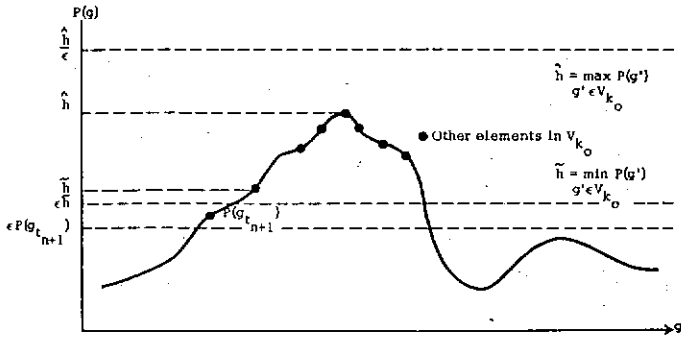


Fig. 5. Geometric illustration of how probability is used in the clustering procedure for a one-dimensional measurement space.

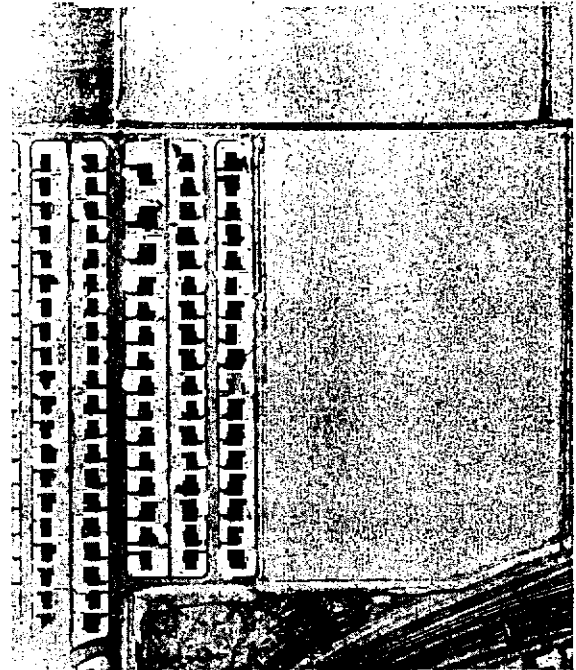


Fig. 6. Image taken over Phoenix, Ariz., with a bandwidth of 400 to 500 millimicrons.

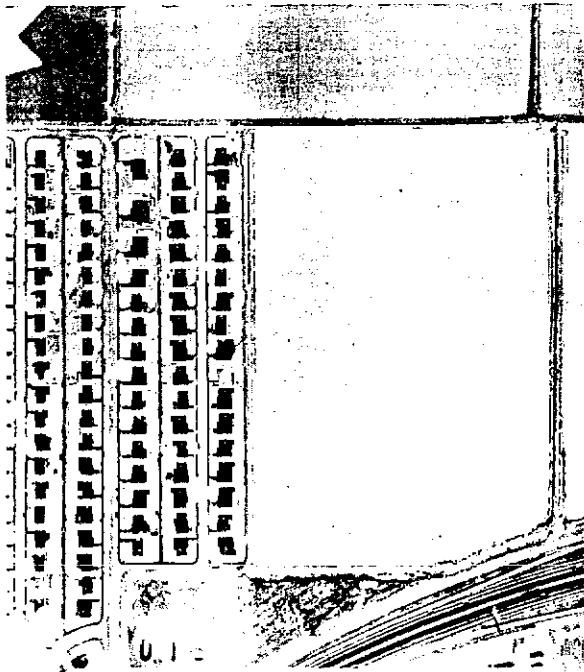


Fig. 7. Image taken over Phoenix, Ariz., with a bandwidth of 520 to 550 millimicrons.

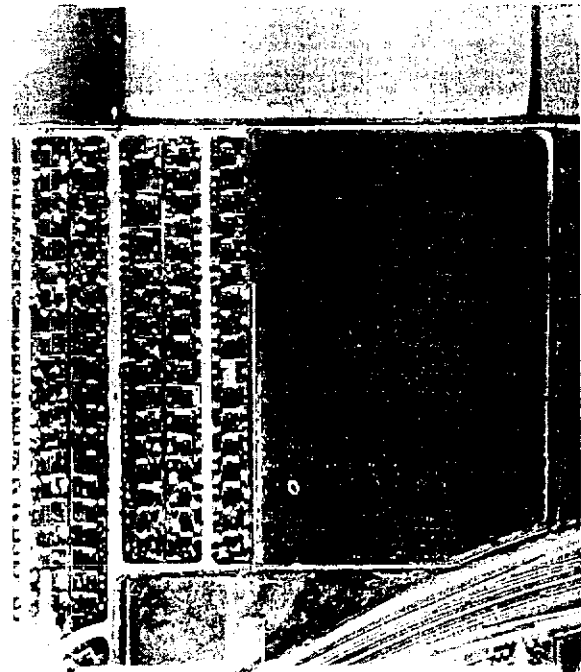


Fig. 8. Image taken over Phoenix, Ariz., with a bandwidth of 810 to 900 millimicrons.

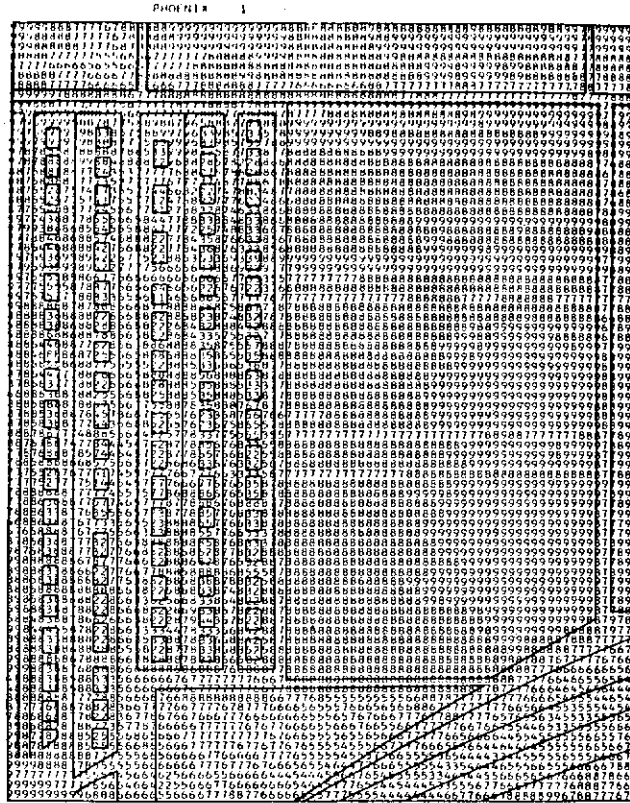


Fig. 9. A 400- to 500-millimicron image quantized to 10 levels.

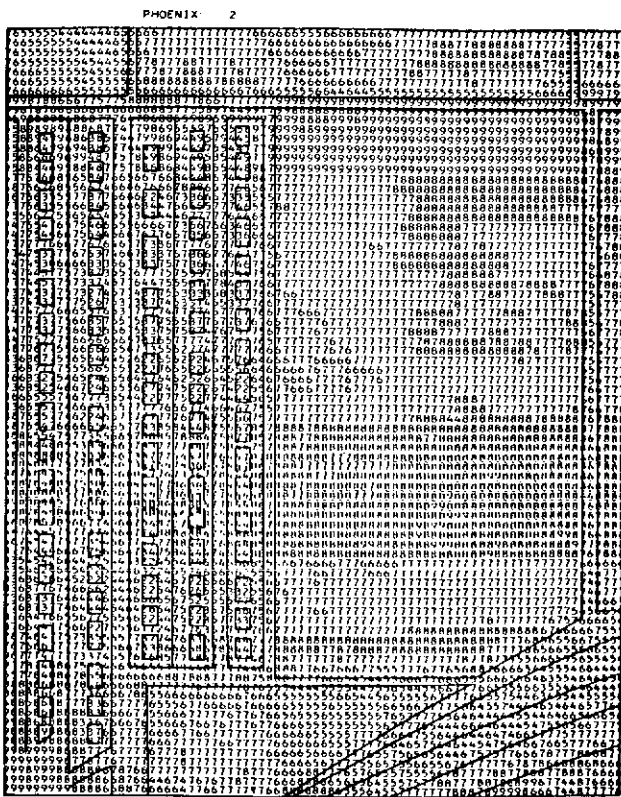


Fig. 10. A 520- to 550-millimicron image quantized to 10 levels.

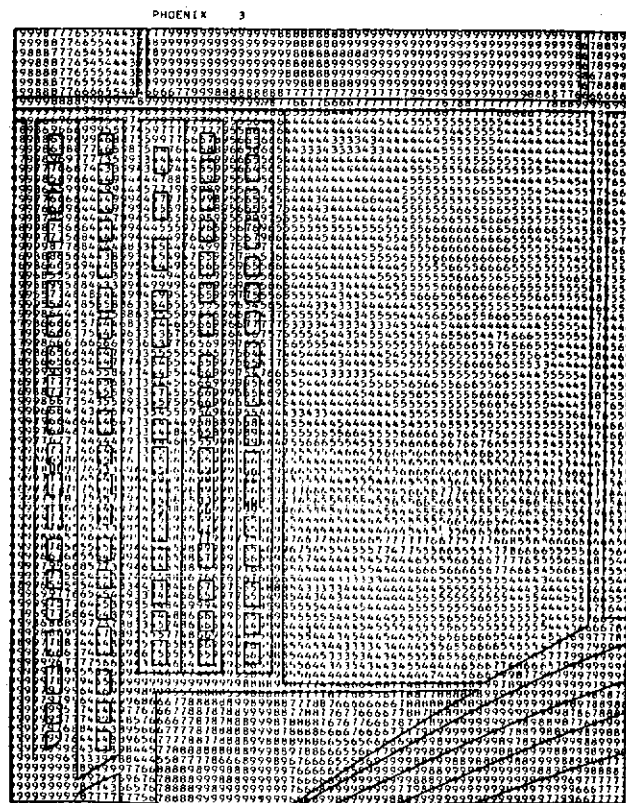
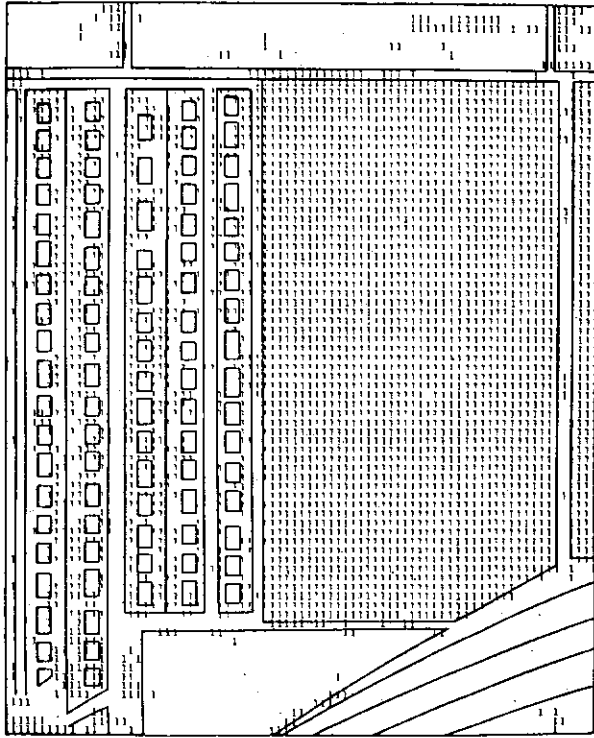
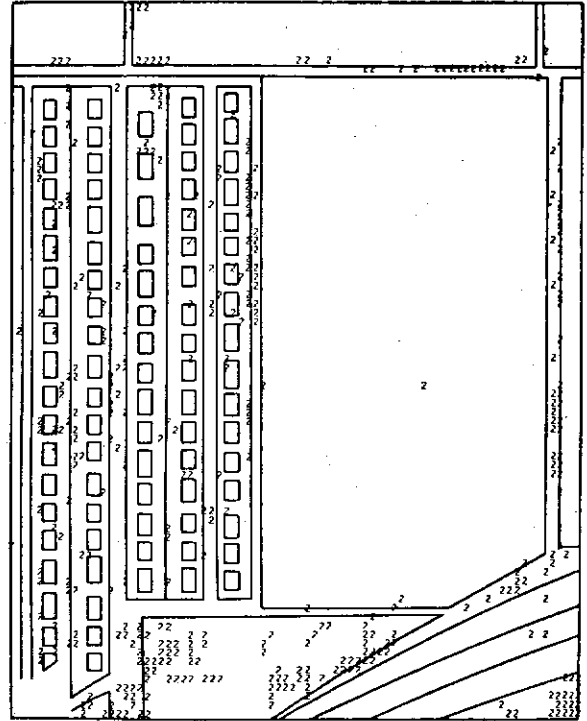


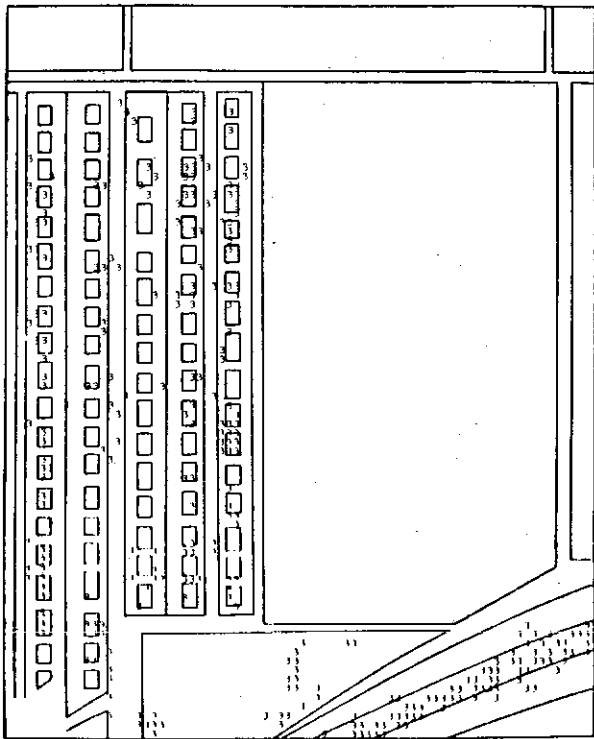
Fig. 11. An 810- to 900-millimicron image quantized to 10 levels.



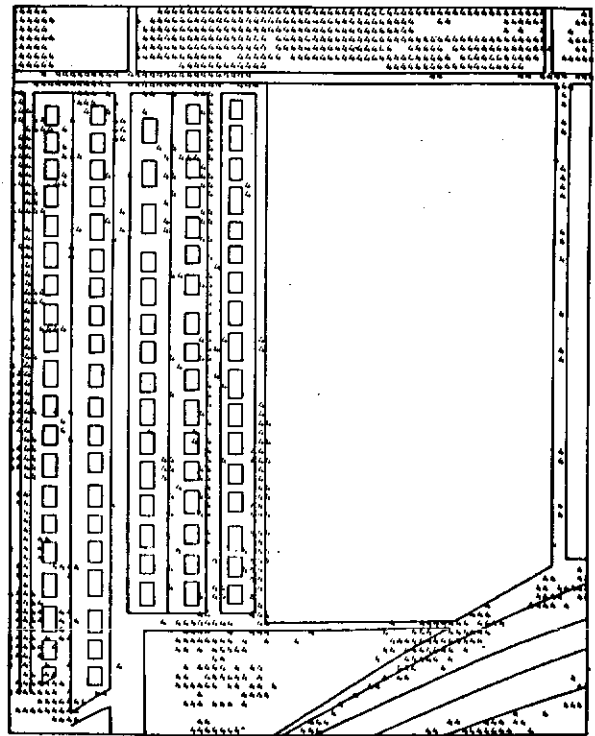
(a) Category 1 identified. The large field and the lawn grass are seen as similar.



(b) Category 2 identified. It seems mainly to be dirt or dirt overgrown with weeds.

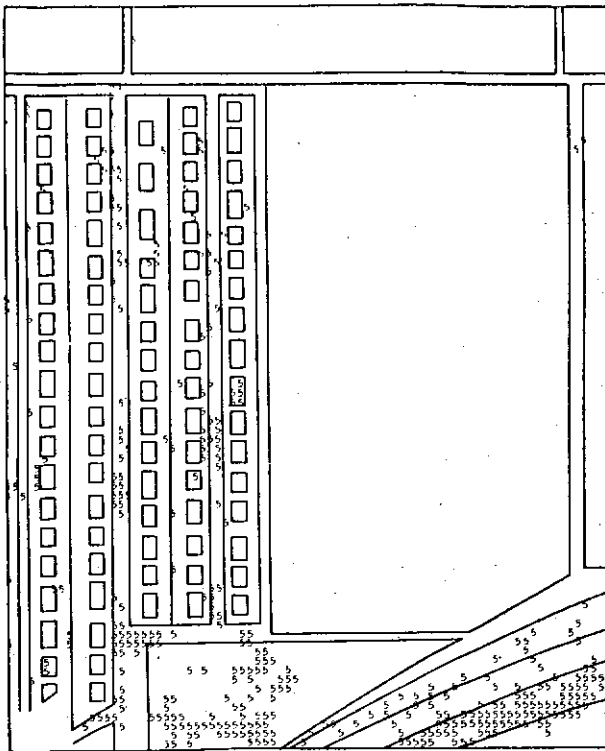


(c) Category 3 identified. The middle part of the turnpike complex is different from the rest. Notice that a few houses are seen as similar to a section of the turnpike. These houses probably had roofs made of material with the same type of reflectance properties as this part of the turnpike.

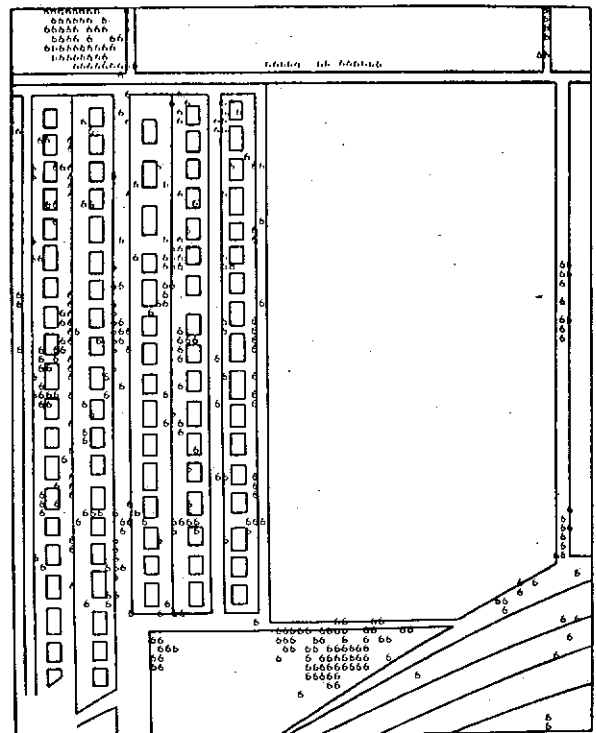


(d) Category 4 identified. It largely consists of the rectangular field at the top.

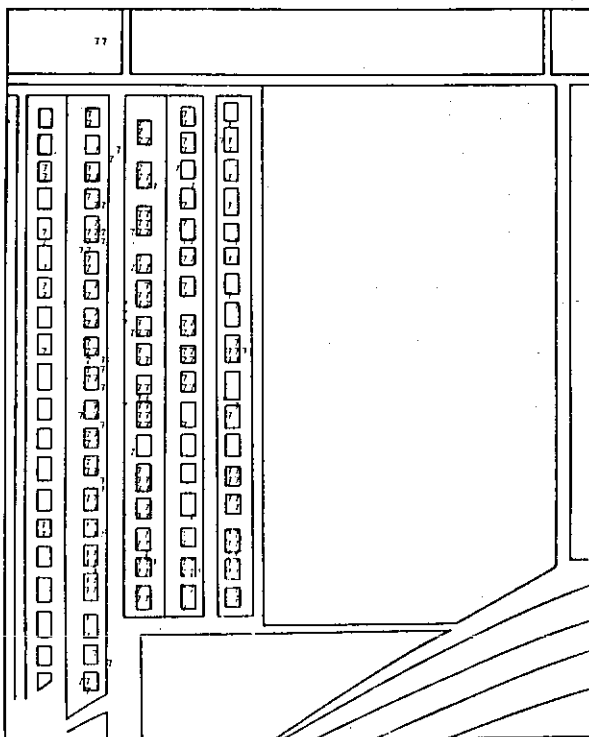
Fig. 12.



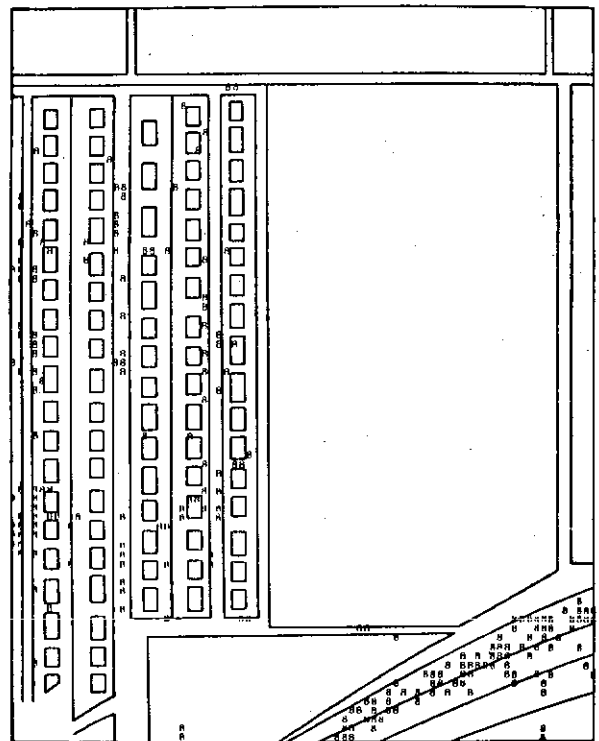
(e) Category 5 identified. It is largely the bottom part of the turnpike. Notice the house which is seen as similar to it. The roofing material on this house must have reflectance properties similar to this section of the turnpike.



(f) Category 6 identified. The uppermost left field is seen as similar to the right part of the triangular wedge-shaped field of dirt overgrown with weeds at the bottom. Perhaps this part of the wedge-shaped field was wetter or drier than the rest of it.



(g) Category 7 identified. Due to the crudeness of the quantization, the grossness of the resolution, and inaccuracies of perhaps one resolution cell in congruencing the images, not all the houses (which are only three or four resolution cells) were seen as similar. However, Category 7 represents a substantial portion of the houses.



(h) Category 8 identified. It consists largely of the upper portion of turnpike complex. It is interesting to note that there are probably three different types of road materials used in the turnpike construction.

enough to V_{k_0} , then we add that g into V_{k_0} . "Similar enough" is determined by the following four conditions:

- 1) $P(g) \geq P(g')$ for every $g' \in S$
- 2) $P(g) \geq \epsilon P(g_{i_{n+1}})$
- 3) $P(g) \geq \epsilon \min_{g' \in V_{k_0}} P(g')$
- 4) $P(g) \leq \frac{1}{\epsilon} \max_{g' \in V_{k_0}} P(g')$.

If there exists a $g \in S$ satisfying these conditions, then we set $V_{k_0+1} = V_{k_0} \cup \{g\}$; if no such g exists, then the clustering procedure is terminated and we set $H_{n+1} = V_{k_0}$.

It is useful to discuss the interpretation of these conditions. Condition 1) means that of the not-yet-clustered elements within a distance 1 from V_{k_0} , we wish to consider only those of highest probability. Condition 2) means that we wish to consider only those elements whose probability is not too much smaller than the center element $g_{i_{n+1}}$ in H_{n+1} . This guarantees that there will not be too wide a variance for the elements in H_{n+1} , and it prevents us from considering background noise. Conditions 3) and 4) restrict our attention to only those elements which have probabilities not too different from those already considered in the cluster. Extreme differences usually indicate an outer boundary for the cell H_{n+1} . Fig. 5 illustrates these conditions geometrically for the case when G is one-dimensional.

Clearly the subsets H_i , $i = 1, \dots, K$, when constructed, are disjoint subsets of G , but it is not certain whether

$$\bigcup_{i=1}^K H_i = G.$$

To make sure the $\{H_i\}_{i=1}^K$ is a partition of G , we perform a nearest-neighbor search procedure.

For each

$$g \in G - \bigcup_{i=1}^K H_i$$

we assign an index as follows: let r be such that

$$\text{dist}(g, H_r) = \text{dist}\left(g, \bigcup_{i=1}^K H_i\right),$$

where we choose the smaller r if r is not unique. After all the points in

$$G - \bigcup_{i=1}^K H_i$$

have been assigned an index, they are included in the corresponding subset H_r . We now have $\{H_i\}_{i=1}^K$, a partition of G with K cells.

V. APPLICATION TO MULTISPECTRAL IMAGERY

Figs. 6, 7, and 8 illustrate three images taken on 70-mm film from a nine-lens multispectral camera. The images are over a suburb of Phoenix, Ariz. Each image recorded the

energy in a specific bandwidth. These were 400 to 500 millimicrons, 520 to 550 millimicrons, and 810 to 900 millimicrons, respectively. Each image was put on an 80×80 grid, and the average density for each resolution cell was measured by hand with a microdensitometer. The densities on each image were then quantized to ten levels; the reduced quantized images are illustrated in Figs. 9, 10, and 11.

The measurement space clustering classification method described in Section IV was programmed for a GE-625 computer and was tried out for the Phoenix imagery. The maximum number of categories K was set equal to 8 while the parameter ϵ was allowed to vary.

An identification of each resolution cell on the quantized imagery was made with respect to the similarity sets or clusters constructed. For small ϵ , less than 0.01, the first couple of similarity sets constructed were so large that they alone covered the entire image. For large ϵ , greater than 0.6, the similarity sets found seemed to have no correspondence to reality—turnpikes, roads, and fields composed the same similarity set, for example. Variations of ± 10 percent in the final value of ϵ had no effect on the results. A value of ϵ of 0.13 gave the best balance with regard to size of similarity set and correspondence with reality.

When K is too small, the clusters constructed tend to be large, indicating only the most distinct and different similarity sets. When K is too large, some clusters can be empty or indicate unimportant noisy pockets. K should be chosen to be the number of distinct homogeneous categories the investigator thinks it is reasonably possible to obtain from the data. Fig. 12 illustrates the identification resulting from a value of 0.13 for ϵ and 8 for K , and indicates an interpretation for each of the resulting similarity sets.

VI. CONCLUSIONS

We have presented two clustering algorithms for multiple images: one which partitions measurement space and one which partitions the image sequence itself. Both algorithms terminate in a finite number of steps since there are only a finite number of data points to be classified. However, the exact number of steps is variable and depends on the nature of the cluster and the number of resolution cells considered.

To the author's knowledge, all clustering procedures suffer because it is required to specify various parameters. The technique proposed here requires specification of K , the maximum number of categories, and ϵ , the probability cutoff parameter. For a good clustering, various combinations of K and ϵ must be tried. It would be useful if a goodness criterion were formulated so that the partition constructed could be evaluated for each (K, ϵ) combination. Then a search can be done on K and ϵ to find the values which give the best result. However, this poses another problem because average minimum distance or entropy-type criteria are not necessarily the best criteria. More fundamental work needs to be done regarding the nature of clusters and exactly what constitutes a good criterion for judging cluster classifications.

Further research can be done in using this type of clustering as a possible coding scheme in transmitting remotely sensed images of planets. Instead of transmitting the measurement space coordinates for each resolution cell, only a code for the cluster in which the resolution cell or the measurement space coordinates reside need be transmitted. This can offer a considerable reduction in bandwidth when used properly.

REFERENCES

- [1] S. S. Wilks, *Mathematical Statistics*. New York: Wiley, 1963, p. 508.
- [2] L. Weiss, "Statistical decision theory in engineering," in *Information and Decision Processes*, R. E. Mochol, Ed. New York: McGraw-Hill, 1960, pp. 170-177.
- [3] R. R. Sokal and P. H. A. Sneath, *Principals of Numerical Taxonomy*. San Francisco: Freeman, 1963.
- [4] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas, Lawrence, Science Bulletin*, March 20, 1958.
- [5] R. L. Mattson and J. E. Rammann, "A technique for determining and coding subclasses in pattern recognition problems," *IBM J. Research Develop.*, July 1965.
- [6] D. B. Cooper and P. W. Cooper, "Nonsupervised adaptive signal detection and pattern recognition," *Information and Control*, vol. 7, no. 3, 1964.
- [7] G. H. Ball and D. J. Hall, "ISODATA, a novel method of data analysis and pattern classification," *Stanford Research Inst., Menlo Park, Calif.*, 1965.
- [8] —, "A clustering technique for summarizing multivariate data," *Behavioral Science*, vol. 12, no. 2, 1967.
- [9] G. S. Sebestyen, "Pattern recognition by an adaptive process of sample set construction," *IRE Trans. Information Theory*, vol. IT-8, pp. 82-91, September 1962.
- [10] G. S. Sebestyen and J. Edie, "Pattern recognition research," *Air Force Cambridge Research Lab., Bedford, Mass., Rept. 64-821 (AD 603 692)*, 1964.
- [11] G. H. Ball, "Data analysis in the social sciences: What about the details?" *1965 Proc. Fall Joint Computer Conf.*
- [12] H. P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *J. Am. Stat. Assoc.*, vol. 62, December 1967.
- [13] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, New York: Hafner, 1967, p. 539.
- [14] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*. New York: Springer-Verlag, 1965, p. 59.

6658
 711074327
 1-361

Reprinted from the PROCEEDINGS OF THE IEEE
 VOL. 57, NO. 4, APRIL, 1969
 pp. 654-665

COPYRIGHT © 1969—THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.
 PRINTED IN THE U.S.A.