COMMUN. STATIST.-THEOR. METH., A5(12), 1163-1191 (1976)

THE TABLE LOOK-UP RULE

Robert M. Haralick

Department of Electrical Engineering &
Remote Sensing Laboratory
The University of Kansas
Lawrence, Kansas 66045

Key Words & Phrases: decision rule; confusion matrix; Bayesian decision rules; Bayes rules; crop determination; binary tables; double hashing technique.

ABSTRACT

The table look-up rule problem can be described by the question: what is a good way for the table to represent the decision regions in the N-dimensional measurement space. This paper describes a quickly implementable table look-up rule based on Ashby's representation of sets in his constraint analysis. A decision region for category c in the N-dimensional measurement space is considered to be the intersection of the inverse projections of the decision regions determined for category c by Bayes rules in smaller dimensional projection spaces. Error bounds for this composite decision rule are derived: any entry in the confusion matrix for the composite decision rule is bounded above by the minimum of that entry taken over all the confusion matrices of the Bayes decision rules in the smaller dimensional projection spaces.

On simulated Gaussian Data, probability of error with the table look-up rule is comparable to the optimum Bayes rule.

1. INTRODUCTION

In the simple Bayes approach to pattern discrimination, a pattern measurement d is assigned to category c* only if

$$P(c^*|d) \geq P(c|d) \text{ for every } c\varepsilon C.$$

There are two distinct ways of implementing this assignment process: in the usual case, we take the pattern measurements to be vectors and for each category c, we estimate the conditional density P(d|c) assuming a convenient multivariate form for P(d|c). When a measurement d arrives for assignment, we plug it into the formula for P(d|c) and assign it to category c* where

$$P(d|c^*)P(c^*) \geq P(d|c)P(c), \text{ for every } c\varepsilon C.$$

The only memory storage needed for this implementation process is for the parameters (mean and covariance) for each density. However, since a density must be computed each time a measurement needs to be assigned, the implementation tends to be compute-bound. This is a serious disadvantage for pattern discrimination using remotely sensed data because the number of measurements tends to be so high.

The other possible implementation procedure is to store the decision rule itself rather than the densities. Define R(d) to be the category the decision rule assigns to measurement d. For a Bayes rule,

$$R(d) = c^* \text{ if and only if}^{(*)} \; P(c^*|d) \geq P(c|d) \text{ for every } c\varepsilon C.$$

Now when a measurement d arrives for assignment, we use d as an address to the table R and look-up the category assignment. When this method is implemented directly, memory storage is needed

(*) Throughout the rest of this paper the phrase "if and only if" will be abbreviated "iff".

for the entire measurement space. This is a lot of memory especially when the dimension of measurement space gets to be above 4 or 5. Also a lot of computer processing time is needed to set up the table since the decision rule needs to be applied to each possible measurement to determine its assignment. However, since the category assignment is retrieved immediately by only an address calculation, the implementation tends to be fast. This is a clear advantage for pattern discrimination using remotely sensed data.

In this paper we explore the various ways a table look-up rule can be implemented and suggest a new implementation based on Ashby's technique of constraint analysis.

2. THE DIRECT TABLE LOOK-UP RULE

Brooner, Haralick and Dinstein (1971) used a table look-up (discrete Bayes rule) approach on high altitude multiband photography flown over Imperial Valley, California to determine crop types. Their approach to the storage problem was to perform an equal probability quantizing from the original 64 digitized grey levels to ten quantized levels for each of the three bands: green, red, and near infrared. Then after the conditional probabilities were empirically estimated, they used a Bayes rule to assign a category to each of the 10^3 possible quantized vectors in the 3-dimensional measurement space. Those vectors which occurred too few times in the training set for any category were deferred assignment. Figure 1 illustrates the decision regions associated with such a table look-up discrete Bayes decision rule. Notice how the quantized multispectral measurement vectors can be used as an address in the 3-dimensional table to look-up the corresponding category assignment.

The rather direct approach employed by Brooner et al. has the disadvantage of requiring a rather small number of quantized levels. Furthermore, it cannot be used with measurement vectors of dimensions greater than four: for if the number of quantized

1 - Alfalfa
2 - Barley
3 - Safflour
4 - Sugar Beet
5 - Lettuce
6 - Onion
7 - Pasture
8 - Bare Soil

DECISION RULE BOUNDARIES
used by: Discrete Bayes Rule

DATA: Raw, Equal Space Quantized

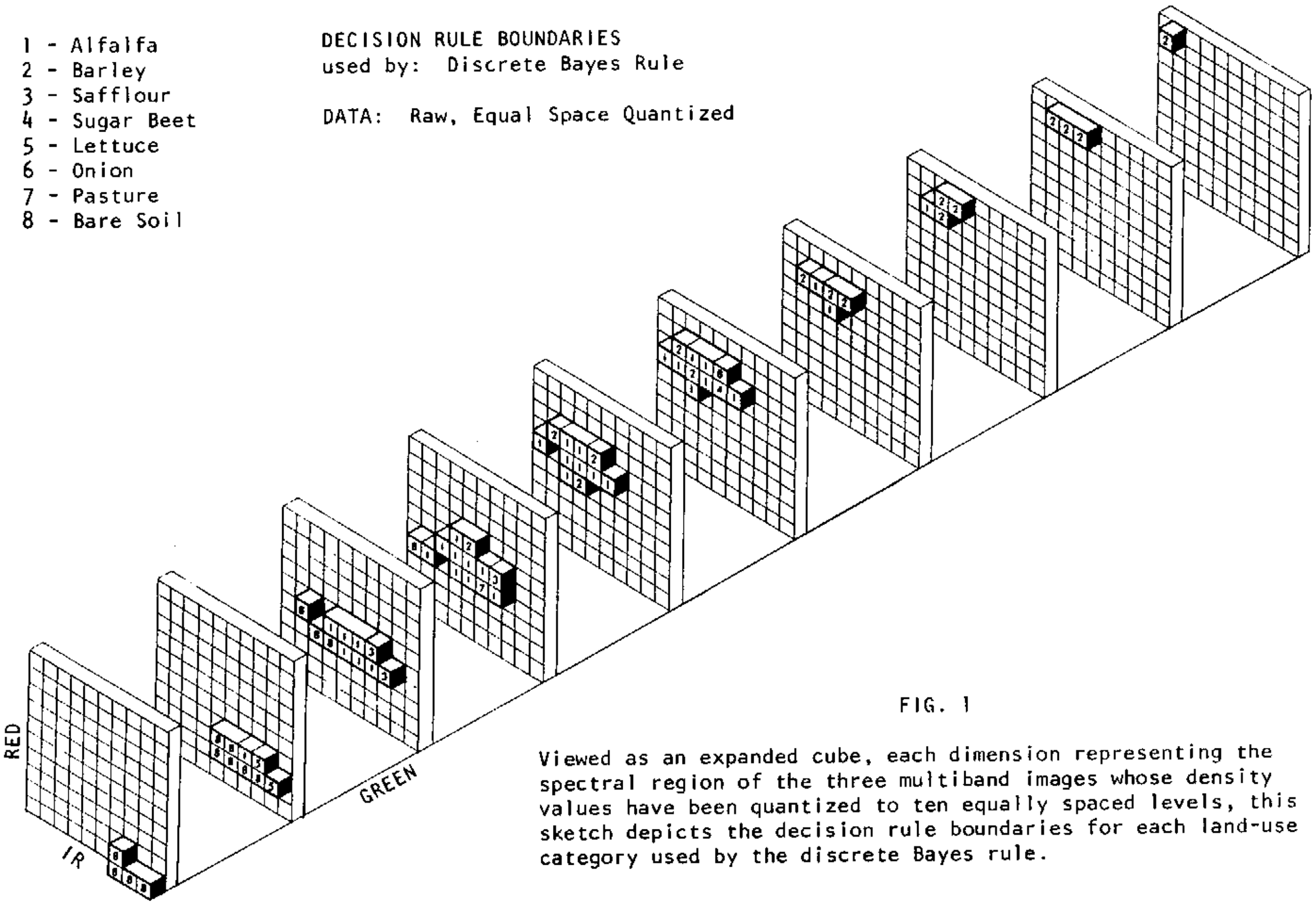


FIG. 1

Viewed as an expanded cube, each dimension representing the spectral region of the three multiband images whose density values have been quantized to ten equally spaced levels, this sketch depicts the decision rule boundaries for each land-use category used by the discrete Bayes rule.

levels is about 10, then the curse of dimensionality forces the number of possible quantized vectors to an unreasonably large size.

Recognizing the grey level precision restriction forced by the quantizing coarsening effect, Eppler, Helmke, and Evans (1971) suggest a way to maintain greater quantizing precision by defining a quantization rule for each category-measurement dimension as follows:

1. fix a category and a measurement dimension component;
2. determine the set of all measurement patterns which would be assigned by the decision rule to the fixed category;
3. examine all the measurement patterns in this set and determine the minimum and maximum grey levels for the fixed measurement component;
4. construct the quantizing rule for the fixed category and measurement dimension pair by dividing the range between the minimum and maximum grey levels into equal spaced quantizing intervals.

This multiple quantizing rule in effect determines for each category a rectangular parallelepiped in measurement space which contains all the measurement patterns assigned to it. Then as shown in Figure 2, the equal interval quantizing lays a grid over the rectangular parallelepiped. Notice how for a fixed number of quantizing levels, the use of multiple quantizing rules in each band allows greater grey level quantizing precision compared to the single quantization rule for each band.

A binary table for each category can be constructed by associating each entry of the table with one corresponding cell in the gridded rectangular parallelepiped. Then define the entry to be a binary 1 if the decision rule assigns a majority of the measurement patterns in the corresponding cell to the specified category; otherwise, define the entry to be a binary 0.

The binary tables are used in the implementation of the multiple quantization rule table look-up in the following way. Order the categories in some meaningful manner such as by prior probability. Quantize the multispectral measurement pattern using

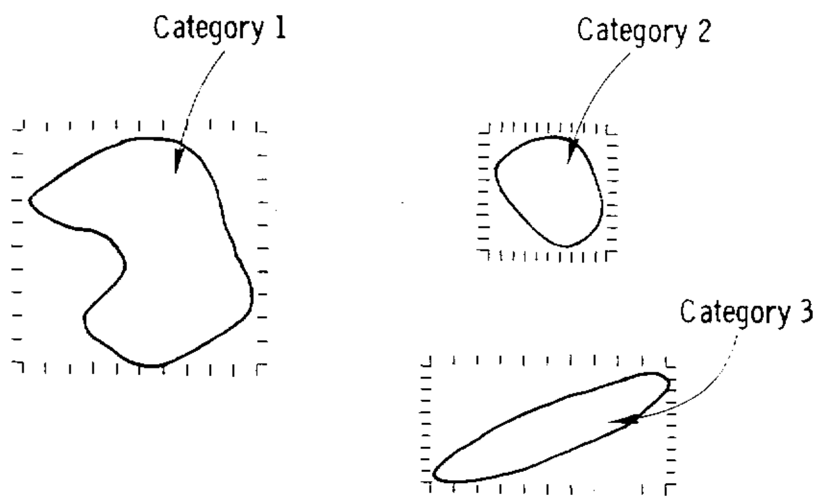


FIG. 2

This illustrates how quantizing can be done differently for each category thereby enabling more accurate classification by the following table look-up rule: (1) quantize the measurement by the quantizing rule for category one, (2) use the quantized measurement as an address in a table and test if the entry is a binary one or binary zero, (3) if it is a binary one assign the measurement to category one; if it is a binary zero, repeat the procedure for category two.

the quantization rule for category c_1. Use the quantized pattern as an address to look up the entry in the binary table for category to determine whether or not the prestored decision rule would assign the pattern to category c_1. If the decision rule makes the assignment to category c_1 the entry would be a binary 1 and, all is finished. If the decision rule does not make the assignment to category c_1, the entry would be a binary 0 and the process would repeat in a similar manner with the quantization rule and table for the next category.

Formally, this kind of table look-up can be described as follows. Let D be measurement space, the set of all possible N-tuple measurements. Let C be the set of categories. For each category $c \varepsilon C$, let D_c be the quantized (discrete and finite) measurement space for category c. Let q_c be the quantizing rule for category c;

$$q_c: \ D \to D_c \ .$$

Note that q_c could quantize some of the components of the N-tuple d to one possible value, in effect excluding that component from consideration.

Let T_c be the decision rule assignment for category c;

$$T_c: \ D_c \to \{0,1\},$$

where

$$T_c(q_c(d)) = 1 \text{ iff } P(c|q_c(d)) \geq P(c'|q_{c'}(d)) \text{ for every } c' \varepsilon C$$
$$= 0 \text{ otherwise.}$$

Then a measurement d is assigned to category c if $T_c(q_c(d)) = 1$.

One advantage to this form of the table look-up decision rule is the flexibility to use different subsets of bands for each category look-up table and thereby take full advantage of the feature selecting capability to define an optimal subset of bands to discriminate one category from all the others. A disadvantage to this form of the table look-up decision rule is the large amount of computational work required to determine the rectangular parallelepipeds for each category and the still large amount of memory storage required (about 5,000 8-bit bytes per category).

Shlien (1975) used a table look-up approach by storing in the table only the category assignments for measurement vectors which requently occur. He used a hashing function to map the measurement vector into the table and reported that if the table is kept at no more than 75% full two distinct vectors are not likely to map to the same table address. Collisions were treated by using the independent double hashing technique described by Amble and Knuth (1974). Shlien indicated that most of the time about 6000 vectors in the table accounted for about 90% of the vectors occurring in an ERTS scene.

3. THE INDIRECT TABLE LOOK-UP RULE

The limitation of the direct approach to the table look-up rule is memory storage. If only some assumptions could be made about the shape of the decision regions or some assumptions about the way a decision region can be represented, or some assumption about the form of the conditional probabilities; perhaps there could be some reduction in storage space associated with the table look-up rule.

Bledsoe and Browning (1959) suggested the following way to approximate the form of the joint probabilities without making a parametric assumption. Let M functions $h_1,\ldots,h_M$ be selected which map the N-dimensional measurement space to smaller K-dimensional discrete and finite feature spaces $F_1,\ldots,F_M$ respectively. Because of the discreteness and small dimensionality of feature space F_m, it is possible to store in tables all the joint probabilities $P_m(c,f)$ of a feature $f \epsilon F_m$ and a category $c \epsilon C$. To assign a category to a measurement d, the M features $h_1(d),\ldots,$ $h_M(d)$ are determined and for each category c, the feature $h_m(d)$ is used as an address to retrieve the probability $P_m(c,h_m(d))$. Then an assignment is made to category c* only if

$$\prod_{m=1}^{M} P_m(c^*,h_m(d)) \geq \prod_{m=1}^{M} P_m(c,h_m(d)) \text{ for every } c \epsilon C.$$

This method is similar to the probability product approximations of Lewis (1959) and the more general product approximations of Ku and Kullback (1969).

Eppler (1974) introduces a table look-up rule which compared to Eppler et. al (1971) enables memory storage to be reduced by five times and decision rule assignment to be decreased by two times. Instead of prestoring in tables a quantized measurement space image of the decision rule, he suggests a systematic way of storing in tables the boundaries or end-points for each region in measurement space satisfying a regularity condition and having all its measurement patterns assigned to the same category.

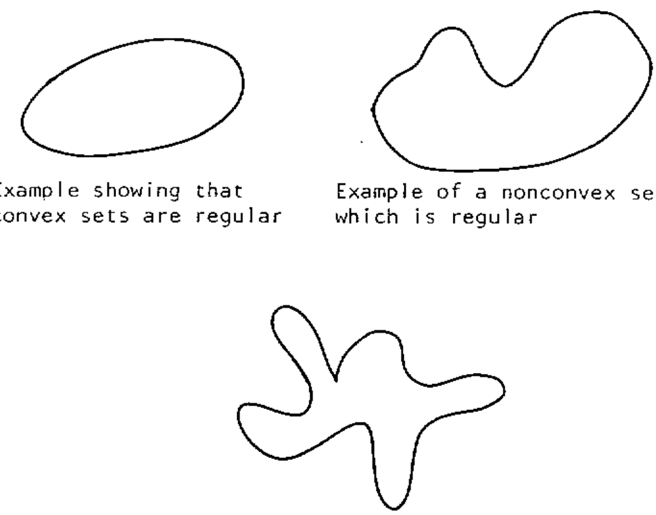
Let $D_q = D_1 \times D_2 \times \ldots \times D_N$ be quantized measurement space. A subset $R \subseteq D_1 \times D_2 \times \ldots \times D_N$ is a regular region iff there exist constants L_1 and H_1 and functions $L_2, L_3, \ldots, L_N, H_2, H_3, \ldots, H_N$, $L_n: D_1 \times D_2 \times \ldots \times D_{n-1} \to (-\infty,\infty)$, $H_n: D_1 \times D_2 \times \ldots \times D_{n-1} \to (-\infty,\infty)$ such that

$$R = \{(x_1,\ldots,x_N) \varepsilon D \mid L_1 \leq x_1 \leq H_1$$
$$L_2(x_1) \leq x_2 \leq H_2(x_1)$$
$$\vdots$$
$$L_N(x_1,x_2,\ldots,x_{N-1}) \leq x_N \leq H_N(x_1,x_2,\ldots,x_{N-1})\}.$$

From the definition of a regular region, it is easy to see how the boundary table look-up decision rule can be implemented. Let $d = (d_1,\ldots, d_N)$ be the measurement pattern to be assigned a category. To determine if d lies within a regular region R associated with category c we look up the numbers L_1 and H_1 and test to see if d_1 lies between L_1 and H_1. If so, we look up the number $L_2(d_1)$ and $H_2(d_1)$ and so on. If all the tests are satisfied, the decision rule can assign measurement pattern d to category c. If one of the tests fails, tests for the regular region corresponding to the next category can be made.

The memory reduction in this kind of table look-up rule is achieved by only storing boundary or end-points of decision regions and the speed-up is obtained by achieving classification most of the time by only using the one or two dimensional tables whose addresses are easier to compute than the three or more dimensional tables required by the direct table look-up decision rule. However, the price paid for these advantages is the regularity condition imposed on the decision regions for each category. This regularity condition is stronger than set connectedness but weaker than set convexity. (See Figure 3.)

Another approach to the table look-up rule can be based on Ashby's (1964) technique of constraint analysis. Ashby suggests



Example showing that convex sets are regular

Example of a nonconvex set which is regular

Example of a nonconvex set which is not regular

FIG. 3

This figure illustrates the relationship between set convexity and regularity.

representing in an approximate way subsets of Cartesian product sets by their projections on various smaller dimensional spaces. Thus, a subset of a Cartesian product set can be approximated by the larger set formed as the intersection of the inverse projections of the projections of the subset onto the smaller dimensional spaces. Using this idea for two-dimensional spaces we can formulate the following kind of table look-up rule.

Let $D_q = D_1 \times D_2 \times \ldots \times D_N$ be quantized measurement space, C be the set of categories, and $J \subseteq \{1,2,\ldots, N\} \times \{1,2,\ldots, N\}$ be an index set for the selected two-dimensional spaces. Let the probability threshold α be given. Let $(i,j) \varepsilon J$; for each $(x_1,x_2) \varepsilon D_i \times D_j$ define the set $S_{ij}(x_1,x_2)$ of categories having the highest conditional probabilities given (x_1,x_2) by

$$S_{ij}(x_1,x_2) = \{c \varepsilon C | P(c|x_1,x2) \geq \alpha_{ij})\} ,$$

where α_{ij} is the largest number which satisfies

$$\sum_{c \varepsilon S_{ij}(x_1,x_2)} P(c|x_1x_2) \geq \alpha \ .$$

Given that components i and j of the measurement pattern take the values (x_1,x_2), $S_{ij}(x_1,x_2)$ is the set of likely categories.

The sets of S_{ij}, $(i,j)\varepsilon J$, can be represented in the computer by tables. In the (i,j)th table S_{ij} the (x_1,x_2)th entry contains the set of all categories of sufficiently high conditional probabilities given the marginal measurements (x_1,x_2) from measurement components i and j, respectively. This set of categories is easily represented by a one word table entry: a set containing categories c_1, c_7, c_9, and c_{12}, for example, would be represented by a word having bits 1, 7, 9, and 12 on and all other bits off.

The decision region R(c) containing the set of all measurement patterns to be assigned to category c can be defined from the S_{ij} sets by

$$R(c) = \{(d_1,d_2,\ldots,d_N)\varepsilon D_1 \times D_2 \times \ldots \times D_N | \\ \{c\} = \bigcap_{(i,j)\in J} S_{ij}(d_i,d_j)\} \ .$$

This kind of a table look-up rule can be implemented by using successive pairs of components (defined by the index set J) of the (quantized) measurement patterns as addresses in the just mentioned two-dimensional tables. The set intersection required by the definition of the decision region R(c) is implemented by taking the Boolean AND of the words obtained from the table look-ups for the measurement to be assigned a category. Note that this Boolean operation makes full use of the natural parallel compute capability the computer has on bits of a word. If the kth bit is the only bit which remains on in the resulting word, then the measurement pattern is assigned to category c_k. If there is more than one bit on or no bits are on, then the measurement pattern is deferred its assignment (reserved decision).

Thus we see that this form of a table look-up rule utilizes a set of "loose" Bayes rules in the lower dimensional projection spaces and intersects the resulting multiple category assignment sets to obtain a category assignment for the measurement pattern in the full measurement space.

Because of the natural effect which the category prior probabilities have on the category assignments produced by a Bayes rule it is possible for a measurement pattern to be the most probable pattern for one category yet be assigned by the Bayes rule to another category having much higher prior probability. This effect will be pronounced in the table look-up rule just described because the elimination of such a category assignment from the set of possible categories by one table look-up will completely eliminate it from consideration because of the Boolean AND or set intersection operation. However, by using an appropriate combination of maximum likelihood and Bayes rule, something can be done about this.

For any pair (i,j) of measurement components, fixed category c, and probability threshold β, we can construct the set of $T_{ij}(c)$ having the most probable pairs of measurement values from components i and j arising from category c. The set $T_{ij}(c)$ is defined by:

$$T_{ij}(c) = \{(x_1,x_2)\varepsilon D_i \times D_j | P(x_1,x_2|c) \geq \beta_{ij}(c)\} ,$$

where $\beta_{ij}(c)$ is the largest number which satisfies

$$\sum_{(x_1,x_2)\varepsilon T_{ij}(c)} P(x_1,x_2|c) = \beta .$$

Tables which can be addressed by (quantized) measurement components can be constructed by combining the S_{ij} and T_{ij} sets. Define $Q_{ij}(x_1,x_2)$ by:

$$Q_{ij}(x_1,x_2) = \{c\varepsilon C| (x_1,x_2)\varepsilon T_{ij}(c)\} \cup S_{ij}(x_1,x_2) .$$

The set $Q_{ij}(x_1,x_2)$ contains all the categories whose respective conditional probabilities given measurement values (x_1,x_2) of components i and j are sufficiently high (a Bayes rule criteria) as well as all those categories whose more probable measurement values for components i and j respectively are (x_i,x_2) (a maximum likelihood criteria). A decision region R(c) containing all the (quantized) measurement patterns can then be defined as before using the Q_{ij} sets:

$$R(c) = \{(d_1,d_2,\ldots,d_N)\varepsilon D_1 \times D_2 \ldots \times D_N | \{c\} = \bigcap_{(i,j)\varepsilon J} Q_{ij}(d_i,d_j)\} .$$

A majority vote version of this kind of table look-up rule can be defined by assigning a measurement to the category most frequently selected in the lower dimensional spaces.

$$R(c) = \Big\{(d_1,d_2,\ldots,d_N)\varepsilon D_1 \times D_2 \times \ldots \times D_N | \\ \# \{(i,j)\varepsilon J | c\varepsilon Q_{ij}(d_i,d_j)\} \geq \# \{(i,j)\varepsilon J | c'\varepsilon Q_{ij}(d_i,d_j) \\ \text{for every } c'\varepsilon C - \{c\}\}\Big\} .$$

3.1 Sequential Table Look-up Rules

The table look-up rule, as other kinds of rules, can also be used in a sequential decision tree procedure in the following way. Each level of the sequential procedure produces a tentative category assignment and the tentative category assignments of level n become an additional dimension of measurement space for layer n+1. Hence measurement space grows by an added dimension each successive level. For all possible distinctions at each level, the sequential procedure is constrained to use the same feature set. Feature sets of different levels, however, can be different.

For each level a feature selection is performed on the measurement space defined for the level in order to determine the optimum measurement space dimensions. The selected measurement space dimensions are then used in a table look-up rule whose category assignments become an added dimension in measurement space for the next level.

Mathematically what happens is this. Let the level 1 decision rule f_1 be an ordinary table look-up rule. Suppose level 1,...level ℓ-1 decision rules $f_1,\ldots,f_{\ell-1}$ have already been defined. Define the level ℓ decision rule in iterative way.

Let $N_{\ell-1}$ be the dimension of measurement space $D_{\ell-1}$ for the (ℓ-1)th level. Define N_ℓ, the dimension of measurement space D_ℓ for the ℓth level by $N_\ell = N_{\ell-1} +1$ and measurement space D_ℓ by

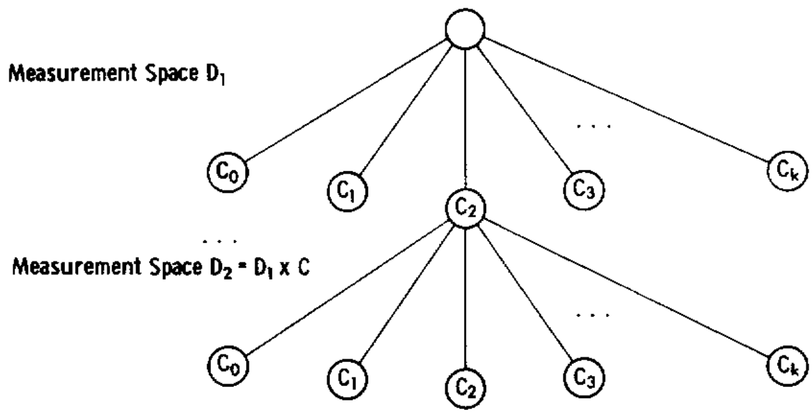
$$D_\ell = \left\{(d_1,\ldots d_{N_{\ell-1}},d_{N_\ell}) \mid (d_1,\ldots,d_{N_{\ell-1}})\varepsilon D\right.$$

and

$$\left.d_{N_\ell} = f_{\ell-1}(d_1,\ldots,d_{N_{\ell-1}})\right\}.$$

The feature selection procedure then uses the measurement space D_ℓ to produce the feature index set $J_\ell \subseteq \{1,\ldots,N\} \times \{1,\ldots,N\}$ as the index set which selects the features. Using a table look-up rule with J_ℓ as described earlier, the decision rule f_ℓ is the determined. Figure 4 shows the decision tree for this sequential rule.

A disadvantage to the sequential rule as just described is that measurement space gets bigger each successive layer. One way of eliminating this problem as well as reducing the computational complexity is to only use successive layers if the assignment of the previous layer reserved judgement. Let the level 1 decision rule f_1 be an ordinary table look-up rule. Suppose level 1,..., level ℓ-1 decision rules $f_1\ldots,f_{\ell-1}$ have already been defined. Empirically determine the category marginal distributions over only those data points assigned reserved judgement by decision rule $f_{\ell-1}$. Use a feature selection procedure on these marginal distributions and determine decision rule f_ℓ using the feature selected components and the category marginal distributions. A data point d is assigned to category c* if and only if for some level $L \geq 1$,



Set of Categories $C = \{C_0, C_1, \ldots C_k\}$, Where C_0 is a Reserved Decision.

FIG. 4

This figure illustrates one way of defining a sequential decision rule by augmenting measurement space by the category classification of the previous layer: The figure shows a 2-layered surface.

$$f_\ell(d) = c_o, \; \ell = 1,\ldots,L-1$$

and

$$f_L(d) = c^* \; .$$

The tree structure for this decision rule is illustrated in Figure 5.

4. MISIDENTIFICATION ERROR BOUNDS

Because the table look-up rule based on tables in a smaller dimensional space than measurement space must necessarily give results which are less optimum than a Bayes rule, it is desirable to determine bounds on the misidentification error. To do this easily we will change our perspective slightly and think of the decision rule in the smaller dimensional space as its induced decision rule in the full measurement space. Ignoring for the moment the relationship between conditional probabilities and the decision rule definition, we will think of a decision rule as a partition in measurement space.

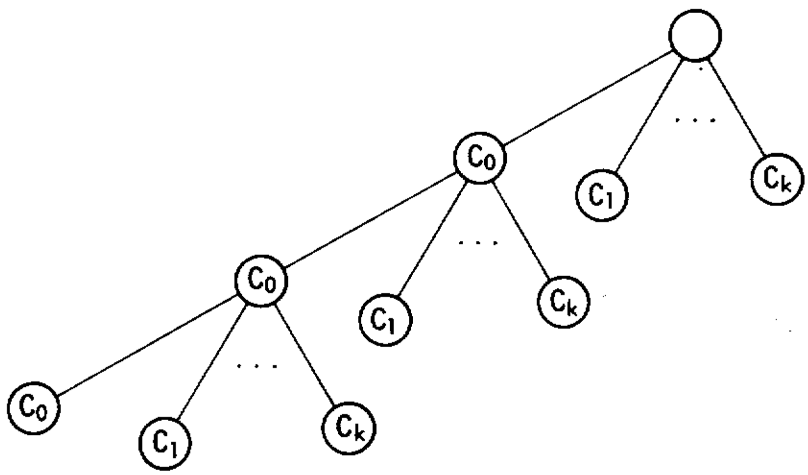


FIG. 5

This figure illustrates the tree structure for the sequential decision rule which uses each successive layer only to resolve, if possible, reserved decisions of the previous layer.

Suppose N different decision rules with no reserved decision regions are determined for measurement space D. Each decision rule can be characterized by the partition generated by its decision regions. Let $\{\Pi_{n1}, \Pi_{n2}, \ldots, \Pi_{nK}\}$ be the partition associated with the nth decision rule. The cell Π_{nk} of the nth partition is the region of all those measurements assigned by the nth decision rule to the kth category.

A composite decision rule (of the table look-up form) can be constructed from the N given decision rules in the following way: a measurement is assigned to the kth category if each and every of the decision rules assigns it to the kth category; if for any measurement a unanimous decision is not possible, then assignment for the measurement is reserved.

Let decision region Π_k of the composite decision rule be the set of all measurements assigned to the k^{th} category and decision region Π_o be the set of all reserved decisions. Then, by definition,

$$\Pi_k = \bigcap_{n=1}^{N} \Pi_{nk}, \quad k=1,2,\ldots,K ,$$

$$\Pi_o = D - \bigcup_{k=1}^{K} \Pi_k .$$

Lemma 1 establishes: (1) upper bounds by category, for the probability of correct identification and the probability of misidentification; (2) lower bounds, by category, for the probability of reserving judgement and for the sum of misidentification probability and reserved judgement probability. The bounds are in terms of the correct identification and error rates in the confusion matrix for each of the N given decision rules.

Let $C = \{c^1, c^2, \ldots, c^K\}$ be the set of categories. Denote by $P_c(c^k)$ the probability of the composite decision rule correctly identifying a unit whose true category identification is c^k, by $P_e(c^k)$ the probability of the composite decision rule misidentifying a unit whose true category identification is c^k, and by $P_r(c^k)$ the probability of the composite decision rule reserving judgement on a unit whose true category identification is c^k. Denote by $P_c^n(c^k)$ the probability of the nth decision rule correctly assigning a unit whose true category identification is c^k, and by $P_e^n(c^k, c^j)$ the probability of the nth decision rule incorrectly assigning a unit whose true category identification is c^k to category c^j. The lemma states:

$$P_c(c^k) \leq \min_{\substack{n \\ n=1,\ldots,N}} P_c^n(c^k) ,$$

$$P_e(c^k) \leq \sum_{\substack{j=1 \\ j\neq k}}^{K} \min_{\substack{n \\ n=1,\ldots,N}} P_e^n(c^k, c^j) ,$$

$$P_e(c^k) + P_r(c^k) \geq \min_{\substack{n \\ n=1,\ldots,N}} \sum_{\substack{j=1 \\ j\neq k}}^{K} P_e^n(c^k, c^j) ,$$

$$P_r(c^k) \geq P(c^k) - \sum_{j=1}^{K} \min_{\substack{n \\ n=1,\ldots,N}} P_e^n(c^k,c^j) .$$

It is also possible to use the error characteristics of a Bayes rule in the smaller dimensional spaces to determine error bounds on the Bayes rule in full measurement space. Lemma 2 gives the upper bound

$$\sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\}$$

for the probability $P_e(c^i:c^j)$ of a Bayes rule confusing categories c^i and c^j. Lemma 3 notes that when measurement space is transformed in any way by a mapping ϕ, then the upper bound of Lemma 2 for the confusion error of the Bayes rule in the transformed space must increase. Lemma 4 states that if $\phi_1,\ldots,\phi_N$ are transformations of measurement space D to spaces $D_1,\ldots,D_N$ respectively, then the error bound of Lemma 2 for the probability of a Bayes rule confusing category c^i and c^j itself can be bounded by

$$\min_{\substack{n \\ n=1,\ldots,N}} \sum_{d_n \varepsilon D_n} \min \{P(d_n,c^i), P(d_n,c^j)\}$$

so that the total probability of error or a Bayes rule in measurement space D can be bounded by

$$\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \min_{\substack{n \\ n=1,\ldots,N}} \sum_{d_n \varepsilon D_n} \min \{P(d_n,c^i), P(d_n,c^j)\} .$$

Lemma 1: Let $\{\Pi_{n1},\ldots,\Pi_{nk}\}$, n=1,...,N be given partitions of measurement space D. Define a new partition $\{\Pi_o, {}_1,\ldots,\Pi_K\}$ by

$$\Pi_k = \bigcap_{n=1}^{N} \Pi_{nk}$$

$$\Pi_o = D - \bigcup_{k=1}^{K} \Pi_k .$$

Then,

$$P_c(c^k) \leq \min_{\substack{n \\ n=1,\ldots,N}} P_c^n(c^k) ,$$

$$P_e(c^k) \leq \sum_{\substack{j=1 \\ j\neq k}}^{K} \min_{\substack{n \\ n=1,\ldots,N}} P_e^n(c^k,c^j) ,$$

$$P_e(c^k) + P_r(c^k) \geq \max_{\substack{n \\ n=1,\ldots,N}} \sum_{\substack{j=1 \\ j\neq k}}^{K} P_e^n(c^k,c^j) ,$$

$$P_r(c^k) \geq P(c^k) - \sum_{j=1}^{K} \min_{\substack{n \\ n-1,\ldots,N}} P_e^n(c^k,c^j) .$$

<u>Proof</u>: By definition,

$$P_c(c^k) = \sum_{d\epsilon\Pi_k} P(d,c^k),$$

$$P_e(c^k) = \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{d\epsilon\Pi_j} P(d,c^k) ,$$

$$P_r(c^k) = \sum_{d\epsilon\Pi_o} P(d,c^k) ,$$

$$P_e^n(c^k,c^j) = \sum_{d\epsilon\Pi_{nj}} P(d,c^k) ,$$

$$P_c^n(c^k) = \sum_{d\epsilon\Pi_{nk}} P(d,c^k) \ .$$

Then,

$$P_c(c^k) = \sum_{d\epsilon\Pi_k} P(d,c^k) = \sum_{d\epsilon \bigcap_{n=1}^{N} \Pi_{nk}} P(d,c^k) \leq \min_{\substack{n \\ n=1,\dots,N}} \sum_{d\epsilon\Pi_{nk}} P(d,c^k)$$

$$\leq \min_{\substack{n \\ n=1,\dots,N}} P_c^n(c^k) \ ;$$

$$P_e(c^k) = \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{d\epsilon\Pi_j} P(d,c^k) = \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{d\epsilon \bigcap_{n=1}^{N} \Pi_{nj}} P(d,c^k)$$

$$\leq \sum_{\substack{j=1 \\ j\neq k}}^{K} \min_{\substack{n \\ n=1,\dots,N}} \sum_{d\epsilon\Pi_{nj}} P(d,c^k) = \sum_{\substack{j=1 \\ j\neq k}}^{K} \min_{\substack{n \\ n=1,\dots,N}} P_e^n(c^k,c^j) \ ;$$

$$P_e(c^k) + P_r(c^k) = \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{d\epsilon\Pi_j} P(d,c^k) + \sum_{d\epsilon\Pi_o} P(d,c^k)$$

$$= \sum_{\substack{j=0 \\ j\neq k}}^{K} \sum_{d\epsilon\Pi_j} P(d,c^k) = \sum_{d\epsilon \bigcap_{\substack{j=0 \\ j\neq k}}^{K} \Pi_j} P(d,c^k)$$

$$\geq P(c^k) - \sum_{j=1}^{K} \min_{\substack{n \\ n=1,\dots,K}} \sum_{d\epsilon\Pi_{nj}} P(d,c^k)$$

$$= P(c^k) - \sum_{j=1}^{K} \min_{\substack{n \\ n=1,\dots,K}} P_e^n(c^k,c^j) \ .$$

Lemma 2: Let $P_e(c^i:c^j)$ be the probability that categories c^i and c^j are confused by a Bayes decision rule in measurement space D. Then

$$P_e(c^i:c^j) \le \sum_{d\varepsilon D} \min\,\{P(d,c^i),\ P(d,c^j)\}\ .$$

Proof: Without loss of generality, we assume that $P(c^i|d) \neq P(c^j|d)$ when $i\neq j$. Let $A_i = \{d\varepsilon D|P(c^i|d) \geq P(c|d)$, for every $c\}$. Since $P_e(c^i:c^j)$ is the joint probability that a unit whose true category identification is c^i is assigned to category c^j or a unit whose true category identification is c^j is assigned to category c^i, we must have

$$P_e(c^i:c^j) = \sum_{d\varepsilon A_i} P(d,c^j) + \sum_{d\varepsilon A_j} P(d,c^i)\ .$$

Let $B_{ij} = \{d\varepsilon D|P(c^i|d) > P(c^j|d)$. Certainly, $B_{ij} \geq A_i$ and $B_{ji} \geq A_j$. Hence,

$$P_e(c^i:c^j) \le \sum_{d\varepsilon B_{ij}} P(d,c^j) + \sum_{d\varepsilon B_{ji}} P(d,c^i)\ .$$

Now notice that if $d\varepsilon B_{ij}$, then $P(d,c^j) = \min\,\{P(d,c^i), P(d,c^j)\}$ and if $d\varepsilon B_{ji}$ $P(d,cj) = \min\,\{P(d,c^j), P(d,c^i)\}$. Also notice that $B_{ij} \cup B_{ji} = D$ and $B_{ij} \cap B_{ji} = \emptyset$. Hence,

$$P_e(c^i:c^j) \le \sum_{d\varepsilon B_{ij}} P(d,c^j) + \sum_{d\varepsilon B_{ji}} P(d,c^i)$$

$$= \sum_{d\varepsilon D} \min\,\{P(d,c^i),\ P(d,c^j)\}\ .$$

Lemma 3: Let D be measurement space and $C = \{c^1, \ldots, c^K\}$ be the set of categories. Let a probability function P be given on $D \times C$. Let a mapping $\phi: D \to D'$ be given which induces a probability function on D'. Then,

$$\sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\} \leq \sum_{d' \varepsilon D'} \min \{P(d',c^i), P(d',c^j)\} .$$

Proof: First notice that since ϕ is a mapping, $\{\phi^{-1}(d') | d' \varepsilon D'\}$ is a partition of D. Also, since D is discrete,

$$P(d',c) = \sum_{d \varepsilon \phi^{-1}(d')} P(d,c) .$$

Hence,

$$\begin{aligned} \sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\} &= \sum_{d' \varepsilon D'} \sum_{d \varepsilon \phi^{-1}(d')} \min \{P(d,c^i), P(d,c^j)\} \\ &\leq \sum_{d' \varepsilon D'} \min \left\{ \sum_{d \varepsilon \phi^{-1}(d')} P(d,c^i), \right. \\ &\qquad \left. \sum_{d \varepsilon \phi^{-1}(d')} P(d,c^j) \right\} \\ &\leq \sum_{d' \varepsilon D'} \min \{P(d',c^i), P(d',c^j)\} . \end{aligned}$$

Lemma 4: Let D be measurement space and $C = \{c^1, \ldots, c^K\}$ be the set of categories. Let a probability function P be given on $D \times C$. Let N mappings $\phi_n: D \to D_n$, $n=1,2,\ldots N$ be given. Then an upper bound on the probability of error, P_e, for a Bayes rule in D can be given by:

$$P_e \leq \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \min_{\substack{n \\ n=1,\ldots,N}} \sum_{d_n \varepsilon D_n} \min \{P(d_n,c^i), P(d_n,c^j)\} .$$

Proof: By a previous lemma, we know that for each mapping ϕ_n,

$$\sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\} \leq \sum_{d_n \varepsilon D_n} \min \{P(d_n,c^i), P(d_n,c^j)\} .$$

Certainly it must then be true for the mapping giving the smallest right-hand side. Hence,

$$\sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\} \leq \min_{\substack{n \\ n=1,\ldots,N}} \sum_{d_n \varepsilon D_n} \min \{P(d_n,c^i), P(d_n,c^j)\} .$$

Now by definition,

$$P_e = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} P_e(c^i:c^j) ,$$

where $P_e(c^i:c^j)$ is the probability that a Bayes rule will confuse categories c^i and c^j. By a previous lemma, $P_e(c^i:c^j)$ is bounded above by

$$\sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\} .$$

Therefore,

$$P_e = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} P_e(c^i:c^j) \leq \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \sum_{d \varepsilon D} \min \{P(d,c^i), P(d,c^j)\}$$

$$\leq \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \min_{\substack{n \\ n=1,\ldots,N}} \sum_{d_n \varepsilon D_n} \{P(d_n,c^i), P(d_n,c^j)\} .$$

5. RESULTS

To illustrate the table look-up technique, we prepared a simulated data set consisting of two normally distributed categories each with 1250 3-dimensional vectors whose components were integer-valued between 0 and 31. The mean vectors were

$$M_1 = \begin{pmatrix} 10 \\ 18 \\ 12 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 14 \\ 14 \\ 18 \end{pmatrix}$$

for categories one and two, respectively.

Each category had the same covariance matrix

$$\Sigma = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 5 & 1 \\ 2 & 1 & 4 \end{pmatrix}$$

Making the actual mean and covariance matrices exactly as required when the data vectors are integer valued is not that easy. The actual mean vectors were

$$M_1 = \begin{pmatrix} 10.03 \\ 17.97 \\ 12.00 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 14.03 \\ 14.01 \\ 18.00 \end{pmatrix}$$

Category one had covariance matrix

$$\begin{pmatrix} 4.1312 & .9895 & 1.9529 \\ .9895 & 5.0481 & .9287 \\ 1.9529 & .9287 & 4.1064 \end{pmatrix}$$

Category two had covariance matrix

$$\begin{pmatrix} 4.1235 & 1.0281 & 2.0432 \\ 1.0281 & 5.2657 & 1.0966 \\ 2.0432 & 1.0966 & 4.1791 \end{pmatrix}$$

Figure 6 shows range plots of category one against category two. Each plot shows the actual range of the central 95% of the probability.

The Mahalanobis distance between the categories is

$$r_o^2 = (M_1 - M_2) \Sigma^{-1} (M_1 - M_2) = 16.2143 .$$

BAND 1

```
CAT01   1-----------------------------------------1
CAT02                     1-----------------------------------------1

        11111111111111111111111111111111111111111111111111111111111
         1                 1                  1                  1
        5.609             9.469              13.531             17.594
```

BAND 2

```
CAT02   1-------------------------------------------1
CAT01                    1-------------------------------------------1

        11111111111111111111111111111111111111111111111111111111111
         1                 1                  1                  1
        8.703             13.156             17.844             22.531
```

BAND 3

```
CAT01   1-------------------------------------1
CAT02                            1----------------------------------1

        11111111111111111111111111111111111111111111111111111111111
         1                 1                  1                  1
        6.750             11.500             16.500             21.500
```

FIG. 6

This figure shows range plots of category one with category two for each of the three components of the data vector. 95% of the probability is contained within the specified range.

The probability of correct classification for category one or two is given by Anderson (1958).

$$\int_{-\infty}^{\frac{1}{2}r_o} \frac{1}{\sqrt{2\pi}} e^{-x^2} dx = .9778 .$$

CONTINGENCY TABLE FOR TEST QNT - 4 TEST BY3 - 1

COL = ASSIGN CAT ROW = TRUE CAT

	R DEC	ONE	TWO	TOTAL	#ERR	%ERR	#SD
UNKWN	0	0	0	0	0	0	0
ONE	0	1151	99	1250	99	8	0
TWO	0	87	1163	1250	87	7	0
TOTAL	0	1238	1262	2500	186	7	0
#ERR	0	87	99	186			
%ERR	0	7	8	7			

(a)

CONTINGENCY TABLE FOR TEST QNT - 4 TEST BY4 - 1

COL = ASSIGN CAT ROW = TRUE CAT

	R DEC	ONE	TWO	TOTAL	#ERR	%ERR	#SD
UNKWN	0	0	0	0	0	0	0
ONE	0	1200	50	1250	50	4	0
TWO	0	124	1126	1250	124	10	0
TOTAL	0	1324	1176	2500	174	7	0
#ERR	0	124	50	174			
%ERR	0	9	4	6			

(b)

CONTINGENCY TABLE FOR TEST QNT - 4 TEST BY5 - 1

COL = ASSIGN CAT ROW = TRUE CAT

	R DEC	ONE	TWO	TOTAL	#ERR	%ERR	%SD
UNKWN	0	0	0	0	0	0	0
ONE	0	1212	38	1250	38	3	0
TWO	0	30	1220	1250	30	2	0
TOTAL	0	1242	1258	2500	68	2	0
#ERR	0	30	38	68			
%ERR	0	2	3	2			

(c)

FIG. 7

This figure shows contingency tables for the table look-up rule using individual component pairs (1,2), (1,3), (2,3) respectively with parameters $\alpha = .4$ and $\beta = 0$.

Using the method of Loftsgaarden and Quesenberry (1965) to estimate the required bivariate conditional probability functions we can compute the table look-up rules for $\alpha = .4$ and $\beta = 0$.

Figure 7 shows three contingency tables for the table look-up rule using components (1,2), (1,3) and (2,3) respectively. Figure 8 shows the contingency table for the intersection table look-up rule using all three component pairs.

The correct identification accuracy was 99% which is actually higher than that expected for a Gaussian classifier. However, there were 361 reserved decisions, 14.44%. Notice, that as required by the error bound calculation, the entries in the contingency tables of Figure 8 are less than the entries for the tables of Figure 7.

The real test for any classifier is the results on actual data. We have been successfully using the table look-up classifier regularly for making land use maps from the LANDSAT imagery. We have also found a way of handling the reserve decisions by

CONTINGENCY TABLE FOR TEST QNT - 4 TEST BY1 - 1

COL = ASSIGN CAT ROW = TRUE CAT

	R DEC	ONE	TWO	TOTAL	#ERR	%ERR	#SD
UNKWN	0	0	0	0	0	0	0
ONE	168	1082	0	1250	0	0	0
TWO	193	6	1051	1250	6	1	0
TOTAL	361	1088	1051	2500	6	0	0
#ERR	0	6	0	6			
%ERR	0	1	0	0			

FIG. 8

This figure shows the contingency table using all three component pairs (1,2), (1,3), (2,3). Note that as required each entry in the table is smaller than or equal to the minimum of the corresponding entries in the contingency tables resulting from classification using the individual component pairs shown in Figure 7.

changing the reserve decision to the category assignment of the nearest resolution cell having a category assignment generated by the table look-up. So for image data, there is a lot of spatial information which can be used to help make good decisions, and the reserve decision inherent in this kind of table look-up is no drawback.

ACKNOWLEDGEMENT

This work was supported by National Science Foundation Grant ENG 74-18981.

BIBLIOGRAPHY

Amble, O. and Knuth, D.E. (1974). Ordered hash table. Computer J. 17, 135-42.

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc.

Ashby, W.R. (1964). Constraint analysis of many dimensional relations. General Systems Yearbook, Vol. 9, 99-105.

Bledsoe, W.W. and Browning, I. (1959). Pattern recognition and reading by machine. Proc. East Joint Computer Conf., 225-32.

Brooner, W.G., Haralick, R.M. and Dinstein, I. (1971). Spectral Parameters affecting automated image interpretation using Bayesian probability techniques. Proc. 7th Int. Symp. Remote Sensing of Environment. Ann Arbor, Michigan: University of Michigan, 1929-49.

Eppler, W.G. (1974). An improved version of the table look-up algorithm for pattern recognition. Proc. 9th Int. Symp. Remote Sensing of Environment. Ann Arbor, Michigan: Environmental Research Institute of Michigan, 793-812.

Eppler, W.G., Helmke, C.A., and Evans, R.H. (1971). Table look-up approach to pattern recognition. Proc. 7th Int. Symp. Remote Sensing of Environment, Ann Arbor, Michigan: University of Michigan, 1415-25.

Ku, H.H. and Kullback, S. (1969). Approximating discrete probability distributions. IEEE Trans. Inf. Theory 15 (4), 444-47.

Lewis, P.M. (1959). An approximating probability distribution to reduce storage requirements. Information and Control 2 (3), 214-25.

Loftsgaarden, D.O. and Quesenberrg, C.P. (1965). A nonparametric estimate of a multivariate density function. Ann. Math. Statist. 38, 1261-65.

Shlien, S. (1975). Practical aspects related to automated classification of LANDSAT imagery using lookup tables. Res. Report 75-2, Canada Centre for Remote Sensing, Department of Energy, Mines and Resources, Ottawa, Canada.