RECOGNITION SYSTEM DESIGN BY STATISTICAL ANALYSIS

L. N. Kanal and N. C. Randall
Philco Corporation
Blue Bell, Pennsylvania

A two-level statistical classification procedure has been applied to the problem of detecting complex targets in aerial photography. At the first level, a set of classification functions designed on the basis of samples from the target class and from other images is used to make subdecisions on local-area statistically-designed features associated with the target class. At the second level these subdecisions are combined into a single decision as to the presence or absence of the target. The nature of the data does not allow for the direct application of classical methods of multivariate discriminant analysis; rather, modifications of classical methods are used. This procedure has been simulated on a digital computer with the aid of a special input-output device which converts imagery to computer language. Excellent results were obtained on independent text samples of actual imagery.

Introduction

Intuitive classification procedures based on concepts of distances and directions, and employing transformations of the coordinate space or projections of the samples along a particular direction have been used from the very beginnings of multivariate statistical analysis. Present classification procedures represent a synthesis of such intuitive concepts with formal probabilistic ideas such as minimizing the probability of misclassification or minimizing the expected loss of misclassification.

The basic assumption underlying probabilistic classification is that there exists for each pattern class a probability distribution $F_g(x_1, x_2, .. x_N)$, where g = 1, 2, .. G, denotes the pattern class or group and the variables X_i, i = 1, 2, .. N represent the observables which are used to characterize the patterns. Members of a group g are considered to be samples from a population which is distributed in an N-dimensional space according to F_g.

The simplest non-trivial situation occurs when there are only two groups whose probability density functions $f_1(x)$ and $f_2(x)$, where $X = (X_1, X_2, .., X_n)$, are known and the patterns to be classified must belong to one of these two groups. For this situation it is well known that optimal classification is obtained by using the likelihood ratio $L(x) = \frac{f_1(x)}{f_2(x)}$. The value of X for a new pattern is substituted in L(x) and the result compared with a threshold. The choice of the threshold depends upon the criterion of optimality being used, the degree of knowledge about the proportion of the two groups in the universe from which patterns are drawn and the costs of correct and incorrect classification.

In practice, the probability distributions are rarely, if ever, completely known. The situation which prevails may be (a) the case of parametric families of distributions, in which functional forms of the distributions are assumed to be known but parameters are estimated from samples; or (b) the non-parametric case in which the only knowledge about the distributions is that which can be inferred from samples and no assumptions are made about the functional forms of the distributions. Expository discussions of the above classification situations with references to the statistical literature and comments on the relationship of this literature to some of the recent work in pattern recognition have been presented previously[1, 2, 3].

In the work described in this paper the traditional assumption, that the form of the probability densities $f_g(x)$ are known, has been used. Specifically, multivariate gaussian functions are assumed. However the estimation and classification procedures differ in two ways from those traditionally used. First, the nature of the data and the relationship of the number of variables to the size of the design sample lead to singular sample covariance matrices. The procedure used for modifying the sample covariance matrices to obtain non-singular estimates of the

population covariance matrices appears more suitable for this problem than the "pseudo-inverse" method. Secondly, classification is performed in two stages rather than in a single stage.

In the following sections of this paper we discuss some of the reasons which led to the procedures being used, the nature of the problem to which they have been applied, and the results which have been obtained on independent test samples.

Multivariate Normal Classification

Estimators are said to be consistent if the estimated value of a parameter approaches the "true" value of the parameter with probability 1 as the sample size is increased indefinitely. The performance of a classification procedure using estimated parameters should be consistent with (i.e., tend to) the optimal performance obtained when true parameter values are used. For classification into one of two groups it has been shown by Fix and Hodges[4] that if $f_1(x, \theta)$ and $f_2(x, \theta)$ satisfy certain weak restrictions as to continuity with respect to θ, where θ denotes the parameters collectively, then using consistent estimates instead of true parameters values in the likelihood ratio procedure will lead to a consistent classification procedure. For the classification of a multivariate observation into one of two groups whose density functions are known except for a number of parameters, if consistent estimators are used for the a priori probabilities and for the unknown parameters, it has been shown by Hoel and Peterson[5] that the Bayes procedure with unknown parameters replaced by estimates is a consistent classification procedure. When sample sizes are finite, and expecially when sample sizes are small, the use of such estimates can only be justified on heuristic grounds. However, it is usually done.

For the case of two normal density functions with the population covariance matrix assumed to be equal for the two groups

$$\text{Log } L(x) = x' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)}) - 1/2\ (\bar{x}^{(1)} + \bar{x}^{(2)})' S^{-1} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

where x' representes the transpose of the column vector $(x_1, x_2, \ldots x_N)$, $\bar{x}^{(g)}$ is the vector of sample means of the g^{th} group and S^{-1} is the inverse of the sample covariance matric S obtained by pooling samples from the two groups. The surface of constant likelihood is thus a hyperplane of the form $\sum_{i=1}^{N} a_i x_i$. When the covariance matrices for the two groups are not assumed to be equal,

$$\text{Log } L(x) = \log \frac{|S_2|^{1/2}}{|S_1|^{1/2}} - 1/2 (x - \bar{x}^{(1)})' S_1^{-1} (x - \bar{x}^{(1)}) + 1/2 (x - \bar{x}^{(2)})' S_2^{-1} (x - \bar{x}^{(2)}).$$

In this case the surface of constant likelihood ratio is no longer a hyperplane but takes the form

$$\sum_{\substack{i=1 \\ j=1}}^{N} a_{i_j} x_i x_j + \sum_{i=1}^{N} b_i x_i .$$

Singular Sample Covariance Matrices

The use of the above classification functions requires the computation of the inverse of the estimated covariance matrix. Regardless of the actual population distribution, if the number of samples available to estimate the covariance matrix is less than or equal to the dimensionality of the measurement space; i.e. if $m \leq N$, the samples will be distributed on a hyperplane of at most (m - 1) dimensions and the estimated covariance matrix will be singular.

If the sample covariance matrix were singular (or near singular) as the sample size m tended to infinity then it would appear reasonable to assume that the population covariance matrix is singular and consider transforming the original set of N observables to a set of $r < N$ linearly independent variables. However the singularity due to the small size of the design sample can hardly be construed as evidence of the singularity of the population covariance matrix.

The approach of restricting the number of observables to a few was justified in the past by citing the complexity and labour involved in computing and inverting sample covariance matrices. In addition to simplifying the computation of statistics dependent on elements of the inverse matrix, this approach usually sufficed to ensure the non-singularity of the sample covariance matrices.

It seems to us that one of the original motivations for the introduction of multivariate statistical techniques was the desire to compensate for the small sample sizes which were available by measuring many characteristics of each member of the sample. Present computational facilities and techniques being considerably beyond those available in the past, one would like to let

the number of observables be determined by the nature of the problem.

If the sample points are distributed on an r dimensional hyperplane r<N, the estimate of the probability density function off this hyperplane is zero. If the population covariance matrices are not assumed to be equal, computational indeterminacy can arise due to the concentration of the estimated probability densities on different subspaces of the measurement space. We then look for a reasonable way of spreading the probability density off the hyperplane such that the estimated multivariate normal density function is non-zero throughout the measurement space. We seek a non-singular estimate of the covariance matrix which approaches the population covariance matrix as the sample size tends to infinity.

Associated with the computational indeterminacy is the conceptual difficulty of assigning class membership to a sample which falls in neither of the subspaces on which the estimated densities are defined.

In some applications the Moore-Penrose Generalized Inverse provides a solution to the singularity problem. Harley[6] has pointed out the unsuitability of the Generalized Inverse solution in the present application. He has also presented a solution which at least in an ad-hoc way resolves both the conceptual and computational difficulties mentioned above. The solution suggested is to add a term proportional to the average variance to each diagonal element of the sample estimate of the covariance matrix. This class of pseudo estimates S_m, can be written as:

$$S_m = f(m, N) E_m + (1 - f(m, N)) \left(\frac{\text{Trace } E_m}{N} \right) U,$$

where E_m is the usual sample estimate of the covariance matrix, $0 \leq f(m, N) \leq 1$ is a monotonically increasing function of m, the number of samples, with $f(m, N) \to 1$ as $m \to \infty$ for N fixed, N being the dimension of the vector of observables; U is the unit matrix. Further discussion of this class of estimates and on the determination of f(m, N) is presented elsewhere[7].

"Two-Layer" Classification

The optimal nature of the likelihood ratio procedure holds as long as the functional forms assumed for the probability densities are valid and the parameters are known or estimated from very large sample sizes. This optimality is no longer assured when parameters are estimated from sample sizes which are not large, or when the underlying distributions deviate from the forms assumed. In practice, both these situations prevail.

Experience indicates that a classification procedure based on dividing the observables into subsets, designing a classification function for each subset and deriving a final classification function which uses the outputs of the first set of statistics, gives better results than a single function based on all the observables.

If we use Q subsets and denote the first set of classification functions by C_j, $j = 1, 2, .., Q$, then the final classification function can be based either on the observables $(C_1, C_2, \ldots, C_Q)$ or on the binary fandom variables obtained by thresholding the C_j; i.e., on $(Y_1, Y_2, \ldots, Y_Q)$, with $Y_j = 1$ if $C_j \geq t_j$ and $Y_j = 0$ (or -1) if $C_j < t_j$. This latter procedure, which combines subdecisions made on subsets of the observables into an overall decision, is what we refer to as "two-layer" classification. Figure 1 shows an example of this procedure. In contrast, the use of a single classification function based on all the observables is referred to as "template-matching."

In previous work of a more deterministic nature, an approach which proved very fruitful was that of making subdecisions on the presence or absence of deterministically designed features and combining these decisions into a final decision as to the object present. The analogy with the present approach leads us to refer to the C_j's, as "statistically designed features."

The concept of "layered" classification in the sense described above has been alluded to in some places in the statistical literature. However, its current emphasis in pattern recognition is a result of the work mentioned above and also the ideas on layered networks introduced in connection with perceptrons and other "trainable" networks. The difference between the C_j's presented here and the first layer functions used in perceptrons is that the latter are restricted to weighting coefficients +1, -1 and 0 and to being linear in the observables. The C_j's, whether linear or quadratic discriminant functions, use coefficients obtained by applying the likelihood ratio procedure to design samples from the alternative classes of patterns.

The rationale for deriving the final classification junction is also different. In the trainable network approach the coefficients of the final linear function are obtained by a sequential adjustment procedure which uses a sample of size one per stage of adjustement. The samples

are cycled through repeatedly and "convergence" of the network is defined as achieving perfect separation of all design samples. In the present approach the rationale for the design of the final classification function is the same as that for the first layer functions. According to this rationale, in theory, the distributions of the binary random variables Y_j, $j = 1, 2, .., Q$, under the alternative classes, are estimated from design samples and used to obtain the final classification function. In practice rather rough approximations to this theoretical approach are necessary when the sizes of available design samples is limited and also because it is necessary to approximate the joint distributions of the binary random variables.

Restricting attention to finite design samples leads to an underestimation of the probability of misclassification. In contrast to the "convergence" criterion used to assess perceptrons, in the present approach performance is assessed by the results obtained on independent test samples.

Local-Area Statistically-Designed Features

The manner in which the observables should be divided into subsets and the number of such subsets which should be used for two-layer classification are questions which cannot be answered uniquely. For the image screening situation, the nature of the problem and the format of the data have suggested the following scheme of subset formation.

The size and quantization of the detection area are determined by the target being sought and the resolution desired. The detection area is divided into a number of subareas. These subareas are chosen in a regular way and are overlapping. The subareas are not designed to encompass particular geometrical features of the target, such as corners, etc.

The classification functions C_j based on these subareas are thus labelled "local-area statistically-designed features' Each subarea is referred to as a "feature block'

The process described above drastically reduces the number of variables which have to be considered at one time. Even with this reduction, however, the available design sample sizes may require, when the C_j, s are quadratic discriminant functions, that the covariance matrices of the two populations be estimated from singular sample covariance matrices.

We note that the two-layer classification method pays a double divident. A single classification function based on all the variables would generally be out of the question for practical reasons; in addition, the two-layer method performs better than a single function.

Preprocessing the Image

Before it is applied to the recognition network, the image is converted into a spatially quantized binary picture. The purpose of this step is (a) to simplify the data for further processing, and (b) to remove the nuisance variables of brightness and contrast. The preprocessor should provide consistent detailed binary pictures of the original image despite wide variations of brightness and contrast level and reversal of contrast. Thus, black objects on a white background will be detected as readily as white objects on a black background.

Several techniques were investigated to determine the most suitable for processing aerial photography. The technique chosen is called Laplacian. The Laplacian function is given by the equation

$$\nabla^2 B = \frac{\partial^2 B}{\partial x^2} + \frac{\partial^2 B}{\partial y^2}$$

where B is the image brightness and x and y are the two image dimensions. Since the function is determined by brightness derivatives rather than by brightness itself, it produces no output on large uniform areas. In the presence of line boundaries or corners, the output is high. Thus the function tends to outline objects and emphasize their contours. The Laplacian filtered image is then thresholded near zero to produce a binary output.

The total system thus consists of three stages of logic: Preprocessing, first level feature detection and final decision. The preprocessing stage uses linear threshold logic whereas the feature detection and final decision stages may use linear or higher order threshold logic.

In practice, Laplacian filtering is obtained by giving the element under consideration a large positive weight, with the surrounding elements having sufficient negative weights to balance the positive weight. Thus, no output is obtained for areas of the picture that have uniform brightness.

The Imagery Screening Problem

The goal of this program is to design a machine that will screen large amounts of tactical aerial photography, enabling the photo-interpreter to spend most of his time making

sophisticated decisions and judgements that require all his skill. This machine must present to the interpreter those photos which have the highest probability of having military significance. While there are many aspects to the imagery screening problem, the most difficult and important is that of automatically detecting tactical targets.

The types of targets that are to be detected include tanks, trucks, aircraft, missiles, artillery, etc. Of course, each photo must be scanned to find these targets in any position or rotation. The detection of these targets must be very reliable, because of (a) the seriousness of missing important targets, and (b) the large number of possibilities of making false alarms.

Tank Detection Experiments

The experimental program had as its goal the detection of M-48 tanks. Toward this end, 100 samples of M-48 tank images were taken from actual tactical photography. The 100 samples represent 18 distinct tanks, each tank having been photographed from several different angles and, in some cases, against different backgrounds due to the tank's motion. In addition, 100 samples of other terrain selected at random were obtained for use as the class from which the tanks are to be separated. Figure 2 shows the imagery that was used. Note that all the tanks are aligned. The ability to detect the tanks in any position and any rotation is discussed in the implementation section.

The Computer Simulation

The imagery in Figure 2 was processed in IMITAC (Image Input to Automatic Computer), a specially constructed scanner which converts images to computer language, and can also convert computer output to images. The Philco IMITAC can scan a 3-by-3 inch image with a 1024 line raster, each line being sampled at 1024 points. An analog-to-digital converter encodes the photographic density to 64 levels. This 6-bit code is then transferred to the UBC (Universal Buffer Controller) of the Philco 2000 computer system, and then to magnetic computer tape.

A set of compatible programs was written for the Philco 2000 computer to process the data from IMITAC. These programs simulate the operations that would occur in an actual screening machine. Some are simply bookkeeping routines, such as the one which separates the data into feature blocks; the MULTINORM program, on the other hand, is based on the statistical classification procedures described earlier.

As generated by the IMITAC equipment, each of the 200 samples, 100 tanks and 100 non-tanks, was in the form of a 32 x 32-element retina. These samples were converted to a one-bit-per-element Laplacian representation. Figure 3 shows a typical Laplacian representation of a tank sample. It is in this form that the statistical procedure must detect tanks. The 32 x 32 retina is then divided into 24 overlapping "feature blocks", each 8 x 8 elements. The position or shape of the feature blocks does not reflect in any way particular features of the tank image. They are simply chosen to cover the entire tank image, occupying a total area of 20 x 28 elements, with 50% overlap between feature blocks. Each feature block is then treated separately until final decisions are arrived at. Each pattern in a feature block is assumed to be a sample from a 64-bit dimensional normal random variable, despite the binary nature of the data.

In order to design a discriminant function for a feature block, the mean vector and covariance matrix of the population must be estimated. To this end, the 100 samples of tanks are divided into two equal groups. One group of 50 samples is used to estimate the population parameters and thereby arrive at a design for the discriminant function. The remaining group of 50 samples is an independent test sample, used to evaluate the performance of the procedure. The same procedure holds for the non-tank samples. In dividing the tank samples, care is taken to insure that none of the 18 distinct tanks fall into both design and test groups.

The design samples are used to estimate the mean vectors and covariance matrices and then the discriminant functions. Both linear and quadratic functions were computed, as previously discussed. The linear discriminant function consists of 64 coefficients or weights, while the quadratic discriminant function has 64 linear weights and 2,016 quadratic weights. In addition, both the discriminant functions were "truncated" by selecting a number of the coefficients which are largest in magnitude and presumably most significant. Thirty-two linear coefficients and 100 quadratic coefficients were selected in each case. This was to test the effect of economizing on the number of weighting elements used.

Each discriminant function was then applied to all of the samples, tank and non-tank, design and test, the result in each case being a number called a "score." The threshold for the feature block is then selected to be halfway between the means of the design scores for tanks

and non-tanks. Each score is compared with the threshold to make a feature decision. This decision says, in effect, "based on this small area of the picture, it appears (or does not appear) that a tank is present."

This procedure is carried through for all 24 feature blocks. In this experiment, the second layer design consisted simply of assigning equal weights to all feature blocks. Thus, the final score is just the number of feature blocks which have made a positive decision, or have "fired." A final decision threshold was not calculated because its value depends heavily on the costs associated with the two types of errors, these costs being determined by the particular application.

The Results

The experimental performance of the statistical classification procedure exceeded all expectations.

The results for a single feature block are plotted in Figure 4 in terms of the cumulative probability distributions of scores for each of the 4 groups of 50 samples. Note that the error rate for the independent test group is only about 6%. Note also that the performance on the design samples far exceeds that on the test samples; no errors would be made using only one feature block! This graphically illustrates the necessity for using independent test samples in evaluating the performance of recognition networks. The smooth curves are normal fits to the observed data.

The final results are shown in Figure 5. Each of the numbers in the blocks represents the number of feature blocks that fired for each sample. Thus, 24 is perfect for tank samples and 0 is perfect for non-tank samples. These results are for the simplest discriminant function tested: the linear truncated. Almost half of the test samples had perfect scores. The tank sample with a score of 11 was examined to determine why it had performed so poorly. It was discovered that this sample was misregistered by a single element, and that by correcting the registry the score increases to 24. The same occurred for the sample with a score of 17 (the reason why misregistration is not a cause for concern is explained in a subsequent section). Therefore, a final threshold could be set anywhere between 7 and 19 without making an error on the 100 test samples. This indicates that good results will be obtained with larger sample sizes.

A number of other experiments were performed with the following general results:

(1) The performance using one-bit Laplacian processed images was about equal to that using the original 6-bit images, and was considerably better than another preprocessing technique.

(2) In terms of separation achieved between scores on design samples, the linear discriminant function is only slightly inferior to the quadratic discriminant function.

(3) The amount of truncation which was used did not greatly affect the performance, particularly for linear discriminant functions.

(4) The performance of the procedure using equal second layer weights was actually superior to a particular attempt to give different weights for each feature block as a function of separability of design scores.

(5) None of the two-layer statistical classification procedures tested made any errors in classifying the 100 independent test samples. This was not the case with deterministic design approaches or random-mask methods.

(6) The performance of the statistical classification procedures far exceeded that obtained using procedures based on the recognition of particular geometric and other features suggested by intuition.

The Implementation

Thus far, we have discussed only the probproblem of detecting targets when they have previously been aligned within the retina. We shall now discuss ways of bringing about this alignment, both in translation and rotation.

First, it is clear that every element of the picture must be searched for the target, since it is impossible to separate potential targets from the background until the detection is actually made. Therefore, an efficient method is needed to scan the photo and apply the discriminant function coefficients to the appropriate elements. This is most efficiently done in the following manner: The photo is scanned by a simple TV type raster. As the elements are scanned, they are serially entered into a shift register or tapped delay line. The discriminant coefficients are realized by resistors attached to

appropriate shift register elements, which are then connected together to perform the summation operation. This is then followed by a threshold circuit which makes the feature subdecision. Imagine that the tank pattern is "frozen" in the fhift register and that the weighting resistors are attached to corresponding elements of the shift register to obtain the best detection of the tank. On reflection, it should be clear that, as the scanning process proceeds, a tank in any translational position in the photo will eventually come into registry with the weighting resistors and be detected. Thus tanks in all translational positions are detected with only a single scan of the photo. Variations in rotation are accommodated by searching at discrete orientations by (a) rotating the TV raster, or (b) designing discriminant functions for each orientation, or a combination of both.

Thus the imagery screening system consists of (a) a high-speed flying spot scanner, (b) a Laplacian preprocessing stage that converts the video to binary data, (c) a shift register correlator with statistically designed coefficients, and (d) the final decision logic. The result is a machine that can rapidly search large amounts of photography and reliably detect a variety of tactical targets in any position.

Concluding Remarks

The work described here and other related efforts lead to a (preliminary) conclusion concerning the relative merit of some competing design approaches. For the identification of small targets in aerial survelliance photography, the two-layer statistical classification method based on local-area statistically designed features and using Laplacian pre-processing is far superior to the "random-mask" methods used (at least in the past) in Perceptrons and other "trainable" networks.

Although the design approach has been tested using independent test samples we feel that the sizes of the design sample and espcially the independent non-tank test sample are considerably smaller than we would like them to be. Work being carried out now will make it possible to test the approach with large sample sizes.

Because the result of the Laplacian preprocessing is to change the grey-scale picture to a a black-and-white picture, the random variables for the first layer of classification junctions are binary. Instead of using the multivariate normal assumption it is possible to consider quadratic and higher order discriminant functions which approximate the likelihood ratio of joint distributions of binary random variables in a different manner. This latter method of approximation is directly related to the binary nature of the variables and in addition avoids the necessity of considering covariance matrices. In the current program such approximations are being investigated.

The sampling error which is obtained when finite sample sizes are used to estimate parameters results in an underestimation of the "true" probability of misclassification. The difference between the probability of misclassification estimated from design samples and that which is actually obtained on independent test samples is therefore of interest. Estimation of these differences and their relationship to sample sizes is currently being investigated.

Acknowledgements

The work described in this paper was performed under Contract DA-36-039-SC-90742 with USAELRDL, Fort Monmouth, New Jersey, Guy Hays, Project Engineer. The authors wish to acknowledge the many contributions made by T. J. Harley, Jr., who was associated with the project from its inception, and J. R. Richards, who did the computer programming. We are also indebted to Dr. P. M. Kelly for his enthusiastic support and encouragement.

References

1. L. Kanal, et al., "Basic Principles of Some Pattern Recognition Systems", Proceedings. National Electronics Conference, Chicago, 1962, p. 279-295.

2. T. J. Harley, et al., Semi-Automatic Imagery Screening Research Study and Experimental Investigation, Philco Corp. Reports 1-6, on Contract DA-36-039-SC-90742, September 1962 to January 1964.

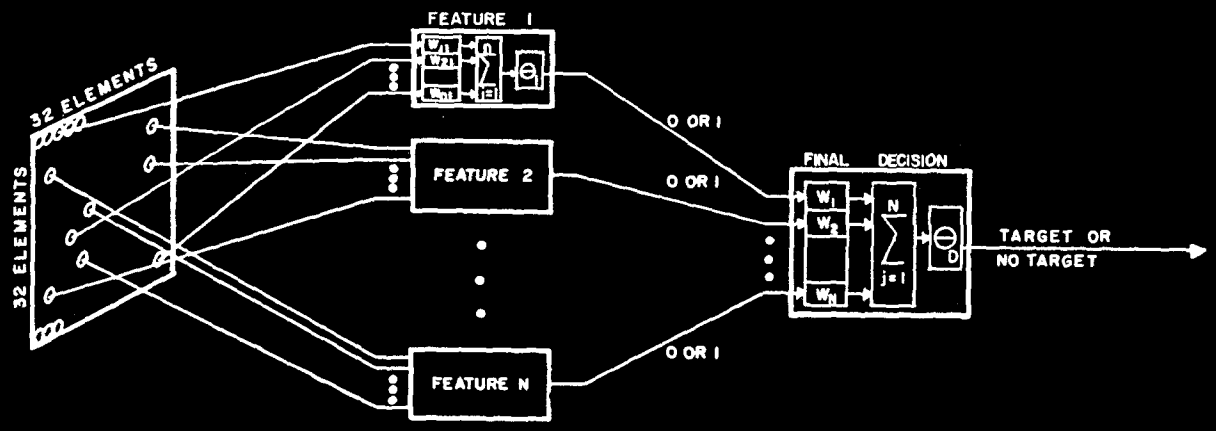
3. L. Kanal and K. K. Nambiar, "On the Application of Discriminant Analysis to Identification in Aerial Photography", Proc. 7th National Convention on Military Electronics, Washington, D.C., September 1963.

4. E. Fix and J. L. Hodges, Discriminatory Analysis: Non-Parametric Discrimination: Consistency Properties, Report #4, USAF School of Aviation Medicine, Randolph AFB, Texas, 1951.

5. P. G. Hoel and R. P. Peterson, "A Solution to the Problem of Optimum Classification", Ann. Math. Stat. Vol. 20, 1949.

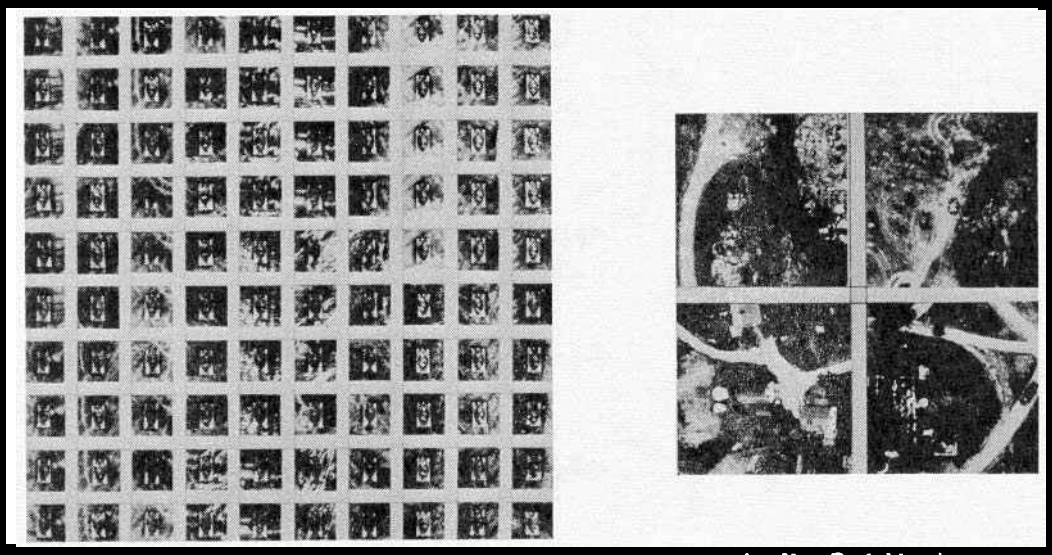
6. T. J. Harley, "A Small Sample Pseudo-estimate of Covariance Matrices", Unpublished Philco Report.

7. Reference 2, Report No. 6, Jan. 1964.



Block Diagram of Detection Logic

Figure 1



a. Tank Mosaic

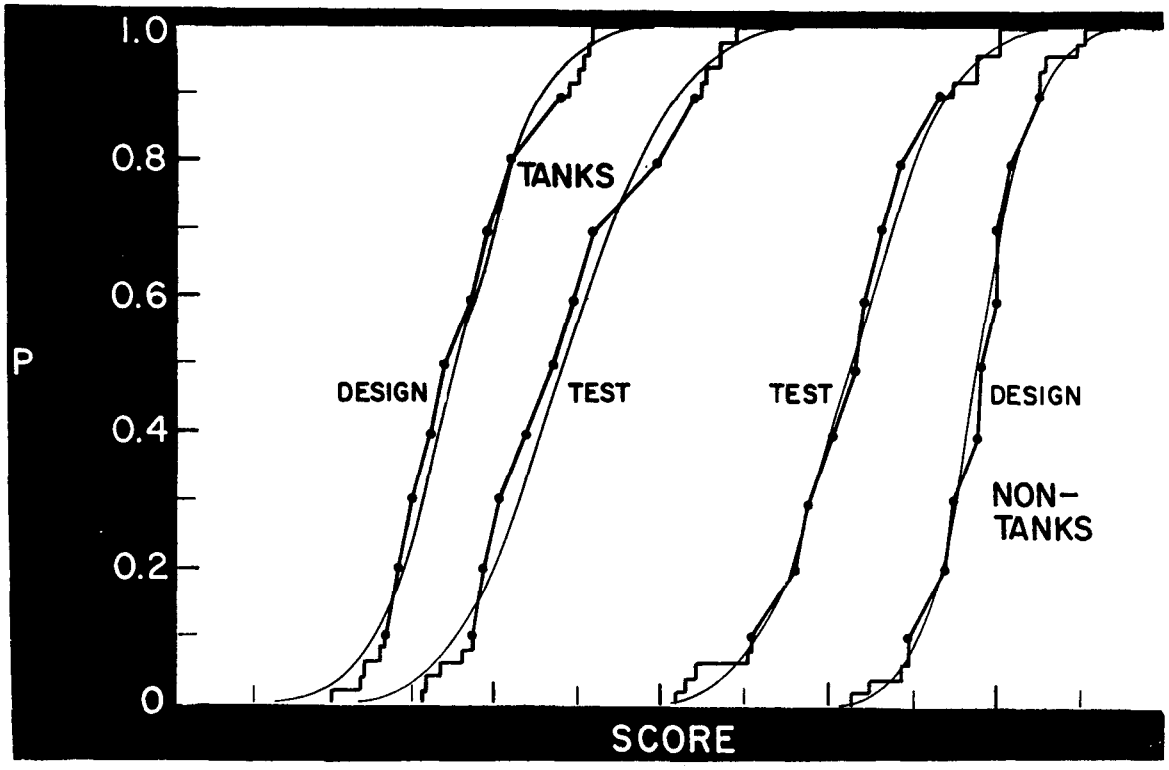
b. Non-Tank Mosaic

Tank and Non-Tank Mosaics

Figure 2

Laplacian Processed Tank Sample

Figure 3



Cumulative Distribution of Scores

Figure 4

22	24	24	23	24	24	24	20	20	24
24	24	24	23	21	24	24	24	22	21
24	24	24	24	24	24	24	24	24	21
24	24	24	24	24	24	24	17	24	20
24	24	24	24	24	23	24	19	24	24
24	24	24	24	21	24	23	22	24	24
21	23	24	24	20	22	23	24	24	24
19	24	20	23	23	20	22	24	24	24
22	24	24	24	24	24	24	24	24	24
24	24	24	24	24	21	(11)	24	24	21

DESIGN TEST

(a) TANKS

0	0	0	0	1	0	0	2	0	3
0	0	1	0	0	0	4	0	0	2
0	0	0	0	0	0	0	0	5	1
0	0	0	0	2	0	0	0	1	2
1	0	1	4	0	1	0	0	(7)	0
2	0	0	0	0	2	3	1	1	0
0	0	0	0	0	1	0	1	0	0
0	0	0	0	0	2	0	1	0	4
0	0	0	0	0	2	3	1	3	1
0	0	0	3	0	1	4	4	0	3

DESIGN TEST

(b) NON-TANKS

Number of Feature Block Detections for Each Sample

Figure 5