

1-1-1977

# A Least-Square Error Approach to LANDSAT Image Classification

Albert Y. Hung

Follow this and additional works at: [http://docs.lib.purdue.edu/lars\\_symp](http://docs.lib.purdue.edu/lars_symp)

---

Hung, Albert Y., "A Least-Square Error Approach to LANDSAT Image Classification" (1977). *LARS Symposia*. Paper 215.  
[http://docs.lib.purdue.edu/lars\\_symp/215](http://docs.lib.purdue.edu/lars_symp/215)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

Reprinted from

**Symposium on  
Machine Processing of  
Remotely Sensed Data**

**June 21 - 23, 1977**

The Laboratory for Applications of  
Remote Sensing

Purdue University  
West Lafayette  
Indiana

IEEE Catalog No.  
77CH1218-7 MPRSD

Copyright © 1977 IEEE  
The Institute of Electrical and Electronics Engineers, Inc.

Copyright © 2004 IEEE. This material is provided with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of the Purdue Research Foundation/University. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

# A LEAST-SQUARE ERROR APPROACH TO LANDSAT IMAGE CLASSIFICATION

ALBERT Y. HUNG  
TRW Defense and Space Systems Group

## I. ABSTRACT

Nonparametric classification methods are often useful in discriminating features or substances even when the functional form of the underlying distributions are unknown to the analyst. One such case is that of geological features, largely devoid of vegetation. Basically, nonparametric classification assumes that there exists a set of discriminant functions (one for each signature) with known functional form except for a set of parameters or weights. In this paper, a nonparametric classifier based on a least-square-error criterion is introduced. Using the designated training samples, an iterative procedure can be formulated which learns the values of the unknown parameters. Consequently, the classification problem is solved by computing the discriminant function and selecting the maximum. Example classifications of LANDSAT MSS scene are studied. Experimental results in the form of thematic maps and percent of correct classification are compared with other well-known techniques such as Bayes and density-slice methods.

## II. INTRODUCTION

Classification of LANDSAT image involves the partitioning of multi-spectral/multi-temporal data vector space into regions defined as signatures or classes. Each picture element (pixel) derived from the MSS imagery will be assigned to a signature identified by a prespecified distinct gray level in the thematic (or classification) map. Basically, these are two different approaches to the classification problem. The parametric approach is characterized by knowing the functional form of class distributions. Thus, the classification problem is treated in the framework of statistic decision theory. The well-known classifiers in this category are Bayes, Eppler, etc. [7, 8, 9]. The nonparametric approach make no probabilistic assumptions. The analyst simply defines the decision boundaries in the n-dimensional data space based on some criterion or similarity measure [2, 10, 11]. In both approaches, if a set of training samples or sites has been used to achieve the decision boundaries, it is called the supervised classification. Otherwise it is called unsupervised classification. The

classifier presented in this paper belongs to the former category. The criterion for data discrimination is the well-known least-square-error approach which has been widely used in pattern recognition [1, 2, 3, 4, 5]. Since the most important task in nonparametric pattern classification is the selection of a set of weights or parameters that defines the discriminant functions, the training method may be viewed as an optimization procedure and the concept of least-square-error can be utilized to form a linear functional.

## III. A LEAST-SQUARE-ERROR IMAGE CLASSIFIER

### A. ALGORITHM

Let  $f_i(x) = W_i^t \phi(x)$ ,  $i=1, 2, \dots, M$ , represent a set of M discriminant functions, where  $\{W_i\}_{i=1}^M$  is a set of M weights (or parameters) to be computed, and  $\phi(x) = (\phi_1(x), \phi_2(x), \dots, \phi_d(x), 1)^t$  are linearly independent, prespecified functions; M is the number of signatures and d is the number of channels or measurements. The image classification problem is solved by computing the discriminant functions and assigning pixel x to signature i if

$$f_i(x) > f_j(x) \quad \forall i, j = 1, 2, \dots, M \text{ and } i \neq j.$$

Now, consider the set of M discriminant functions as a transformation which maps all multidimensional patterns (or data vectors) from signature i to a neighborhood of some M-dimensional fixed vector  $e_i = (e_{i1}, e_{i2}, \dots, e_{iM})^t$ . The mean-square-error criterion is utilized to formulate a linear functional so that the unknown parameters of the transformation can be computed.

$$J = \frac{1}{N} \sum_{k=1}^M \sum_{i=1}^M \sum_{j=1}^{N_i} (x_j^{t(i)} w_k - e_{ik})^2$$

where  $x_j^{t(i)}$  represents the  $j^{\text{th}}$  training sample of signature i

$N_i$  = Number of training samples from signature i



$$D[n] = \Psi W[n] - E[n]$$

$$W[n+1] = W[n] + \Psi^{\#} \Delta E[n]$$

$$E[n+1] = E[n] + \Delta E[n]$$

where

$$\Delta E[n]_{ij} = \rho D[n]_{ij} \quad \text{if } e[n]_{ij} e_j[0] \geq e[n]_{ij} e_l[0] \quad \text{for all } l \neq j$$

$$= 0 \quad \text{otherwise}$$

Here,  $e[n]_{ij}$  denotes the  $i^{\text{th}}$  row of matrix  $E[n]$  corresponding to signature  $j$ . The convergence proof of the recursive scheme is provided in Appendix B.

The properties of least-square-error Criterion and the Bayesian method have been investigated by a number of authors [2, 3, 4, 5]. Patterson and Womach [4] have shown that for two classes pattern classification, the least-square-error approach is equivalent to the Optimal Bayes approach for normally distributed data having identical covariance matrices. Furthermore, Wee [5] proved that the discriminant functions obtained by the generalized inverse approach are closest among all linear functions to Optimum Bayes discriminant functions in a mean-square-error sense as the number of training patterns approaches infinity.

#### B. SOFTWARE IMPLEMENTATION

Figure 1 shows the configuration of the image classification system employed at TRW. The least-square-error classifier consists of two software modules: 1) NTRAIN - for nonparametric training, and 2) NCLASS - for nonparametric classification. NTRAIN designates the training and evaluation sets by making use of the graphics overlay feature of a COMTAL 8000 image display system. This is accomplished by setting the "bits" of the graphic overlay using a track-ball cursor to automatically read CRT pixel addresses. Once an overlay has been defined, its "bits status" can be used to identify the pixel addresses of interest in multitemporal/multi-spectral images stored on a disk. Furthermore, it is possible to have the training graphics on disk and for combination with all graphics to form a joint graphics overlay for later use in the mode of selective classification and/or performance evaluation. Besides computing the unknown parameter matrix  $W$  and generating a parameter file, NTRAIN also calculates the average gray level and pixel scatter matrix for each signature. This piece of information is useful in conducting pixel rejection tests in NCLASS.

NCLASS assigns a unique signature to each pixel to be classified according to the decision rule of equation 1 in part A. The inputs to module NCLASS are the parameter file and the pixel interleaved multispectral data from a specified image source file.

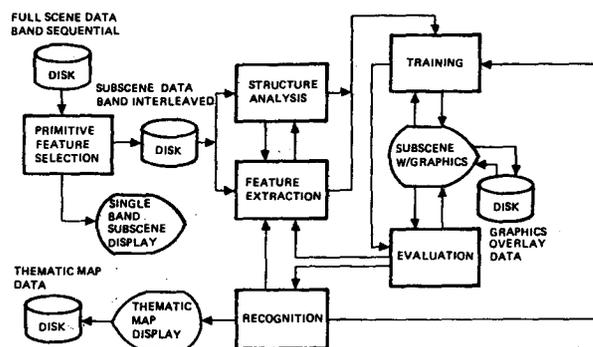


Figure 1. Diagram of the Image Classification System

#### IV. APPLICATION RESULTS

An experimental study of the Least-Square-Error (LSE) classifier was conducted using LANDSAT images (Scene ID 1072-18001) of Goldfield Nevada. The full scene of Goldfield is shown in Figure 2 and the extracted subsce is shown in Figure 3 along with a density-slice thematic map of the area.

The purpose of this experiment was to make a comparison of the results obtained by the use of LSE classifier as well as the well-known Bayes classifier and density-slice methods. In fact, the famous Goldfield test site near the Mud Lake area has been investigated by a number of researchers [12, 13, 14]. In this test, six geological features have been selected for classification as listed in Table 1.

Table 1. Geological Features Selected for Classification

Signature Number	Feature Name	Gray Level Assigned in Classification Map
1	Playa	10
2	Basalt and Vegetation	21
3	Felsic rocks	31
4	Basalt	42
5	Alluvial deposits	52
6	Altered Zone with Limonite	63
7	Unknown	0

For each signature, training was performed non-recursively using a training site designated by cursor positioning on a CRT display. Identical training sites were used for all three classifiers.

In addition, the training graphics literal overlays were combined to form a joint graphics overlay for later use in selective classification and performance evaluation. The percentages of correct classification based on the same training data are tabulated in matrix form as shown in Table 2. Each entry consists of three numbers: the upper number represents the percent of correct classification for the LSE classifier; the middle, for the Bayes classifier; and the bottom, for the density-slicing technique. The ideal result would be a score of 100 percent along the diagonal and zero elsewhere. The actual evaluation matrix indicates that the results of Bayes and LSE classification are similar except in signatures 4 and 5. LSE classifier treats basalt as if it were basalt and vegetation; on the otherhand, the Bayes classifier correctly identifies basalt training samples better than half the time. This result may be explained by the fact that the sample mean vectors extracted from two pairs of training sites (i.e. basalt vs. basalt with vegetation and felsic vs. alluvial soil) show no significant difference in magnitude and direction; however, there is a certain detectable difference in the sample covariance of four spectral bands. The non-recursive LSE classifier employed here generates hyperplanes in decision space, while the Bayes classifier constructs quadratic decision surfaces based on estimated sample covariances.

The performance of the current LSE classifier can be improved by either using a second or higher order  $\phi(x)^*$  function so as to generate the high order decision surfaces or by incorporating the recursive scheme to obtain an optimal E matrix before classification. The classification maps generated by the LSE and Bayes classifiers are shown in Figure 4. Both maps agree well with geologic ground truth, except that the LSE approach tends to put an equal emphasis on basalt and vegetation and felsic rocks as compared to the wide range of felsic rocks of Bayes classification. The density-slice classification map contains a large percentage of unknown class assignments. This is due to the fact that the thresholding technique was implemented without a majority decision rule. Therefore, pixels falling in the overlapping area of the parallelograms are automatically assigned to the null class. The density-slice software can be modified to incorporate a majority decision rule at the cost of processing speed.

#### V. CONCLUSION

The least-square-error classifier has been shown to be a useful tool in LANDSAT image classification. It is superior to the density-slice technique. Application results in Section IV indicate that LSE discrimination can be a useful alternative to the parametric Bayes classification. Furthermore, using high-order discriminant functions and/or recursive

\* Sometimes called a  $\phi$ -machine in pattern recognition. [15]

training method, the LSE classifier can potentially improve the classification performance so that it can consistently out-perform the Bayes approach under nonparametric conditions.

Table 2. Selective Classification Results

		Signature Number					
		1	2	3	4	5	6
Signature Number	1	100	0	0	0	0	1.9
		100	0	0	0	0	0.1
		100	0	0	0	0	0
	2	0	96.7	21.1	79.1	15.5	1.2
		0	84.2	3.8	36.4	0	0
		0	11.4	1.1	4.1	0	0
	3	0	2.9	65.5	7.5	34.0	50.7
		0	4.0	64.0	10.1	8.8	55
		0	0	24.4	0.2	34.0	3.1
	4	0	0	0	12.1	0	0
		0	11.8	9.4	52.3	0	0
		0	0	0.1	1.1	0	0
	5	0	0.4	8.2	0.2	48.2	22.3
		0	0	17.7	0	91.1	27.7
		0	0	8.9	0	11.2	26.1
	6	0	0	5.1	0.8	2.3	23.2
		0	0	1.1	0	0.1	17.0
		0	0	3.9	0	0	30.0
	7	0	0	0.2	0.3	0	0.7
		0	0	4.0	1.2	0.1	0.3
		0	88.6	61.7	94.7	54.8	50.8
	Sample Size	1955	1118	1115	665	920	737

#### APPENDIX A

##### GENERALIZED INVERSE APPROACH FOR UNKNOWN WEIGHTS COMPUTATION

Generalized inverse computation can be used to furnish a quick solution to image classification problem using the least-space-error criterion. Since the rows of matrix E can be interpreted either as reference points or as cost vectors and in both cases it is a predetermined matrix, the minimum of J can be obtained by letting  $\partial J / \partial W = 0$ .

This implies that:

$$W = (\psi^t \psi)^{-1} \psi^t E = \psi^\# E$$

where  $\psi^\#$  is called the generalized inverse.

Let

$$\overline{x[i]} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j[i]$$

$$\overline{xx^t} = \frac{1}{N} \sum_{i=1}^M \sum_{j=1}^{N_i} x_j[i] x_j^t[i]$$

$$\overline{x} = \frac{1}{N} \sum_{i=1}^M N_i \overline{x[i]}$$

then,

$$W = \left( \sum_{i=1}^M \sum_{j=1}^{N_i} x_j[i] x_j^t[i] \right)^{-1} \sum_{i=1}^M \sum_{j=1}^{N_i} x_j[i] e_i^t$$

$$= N^{-1} (\overline{xx^t})^{-1} \left( \sum_{i=1}^M N_i \overline{x[i]} e_i^t \right)$$

Set

$$e_i^t = (C(1/i), C(2/i), \dots, C(M/i)) = c^t(i)$$

and assume

$$c(j/i) = 0 \quad \text{if } i=j$$

$$= c > 0 \quad \text{otherwise}$$

$$W = c(\overline{xx^t})^{-1} \left( \overline{x} - \frac{N_1}{N} \overline{x[1]}, \overline{x} - \frac{N_2}{N} \overline{x[2]}, \dots, \overline{x} - \frac{N_M}{N} \overline{x[M]} \right)$$

A reasonable decision rule is:

Decide  $x \in$  class  $i$  if

$$||x^t W - c^t(i)||^2 < ||x^t W - c^t(j)||^2$$

for all  $j \neq i$

Expanding the above equation, the decision rule becomes:

Decide  $x \in$  class  $i$  if

$$c \sum_{\substack{k=1 \\ k \neq i}}^M x^t W_k - \frac{1}{2}(M-1)c^2 > c \sum_{\substack{k=1 \\ k \neq j}}^M x^t W_k - \frac{1}{2}(M-1)c^2$$

or, if

$$x^t W_j > x^t W_i \quad \text{for all } j \neq i$$

where

$$w_i = c(\overline{xx^t})^{-1} \left( \overline{x} - \frac{N_i}{N} \overline{x[i]} \right)$$

## APPENDIX B CONVERGENCE PROOF

The convergence proof of the recursive algorithm can be divided into two parts.

Part 1. If the constraint on  $E$  is violated, the algorithm will be terminated since  $\delta E[n] = 0$  for all  $n$ .

Part 2. Assume that the constraint on  $E$  holds. It must be shown that the algorithm converges. That is,  $||D[n]|| \rightarrow 0$  as  $n \rightarrow \infty$ .

Two matrix identities will be proved first.

(a)  $\Psi^t D[n]$

$$= \Psi^t (\Psi W[n] - E[n])$$

$$= \Psi^t (\Psi \Psi^{\#} E[n] - E[n])$$

$$= (\Psi^t \Psi \Psi^{\#} - \Psi^t) E[n]$$

$$= (\Psi^t \Psi (\Psi^t \Psi)^{-1} \Psi^t - \Psi^t) E[n] = 0$$

(b) Trace  $\{D[n]^t (\Psi \Psi^{\#} - I)^t (\Psi \Psi^{\#} - I) D[n]\}$

$$= \text{Trace} \{D[n]^t [(\Psi \Psi^{\#})^t (\Psi \Psi^{\#}) - \Psi \Psi^{\#} - \Psi \Psi^{\#} + I] D[n]\}$$

$$= \text{Trace} \{D[n]^t [\Psi (\Psi^t \Psi)^{-1} \Psi^t \Psi (\Psi^t \Psi)^{-1} \Psi^t - 2\Psi \Psi^{\#} + I] D[n]\}$$

$$= \text{Trace} \{D[n]^t (I - \Psi \Psi^{\#}) D[n]\}$$

$$= \text{Trace} \{D[n]^t D[n]\} - \text{Trace} \{D[n]^t \Psi \Psi^{\#} D[n]\}$$

$$= ||D[n]||^2 - \text{Trace} \{(\Psi A[n] - E[n])^t \Psi \Psi^{\#} D[n]\}$$

$$= ||D[n]||^2 - \text{Trace} \{(\Psi \Psi^{\#} E[n] - E[n])^t \Psi \Psi^{\#} D[n]\}$$

$$= ||D[n]||^2 - \text{Trace} \{E[n]^t (\Psi \Psi^{\#} - I)^t \Psi (\Psi^t \Psi)^{-1} \Psi^t D[n]\}$$

$$= ||D[n]||^2 \quad \text{(by matrix identity a).}$$

Now define  $V(D[n]) = ||D[n]||^2$ , a positive definite function.

$\Delta V(D[n])$

$$= V(D[n+1]) - V(D[n])$$

$$= V(\Psi W[n+1] - E[n+1]) - V(D[n])$$

$$= V(\Psi(W[n] + \Psi^{\#} \delta E[n]) - E[n+1]) - V(D[n])$$

$$= V(\Psi W[n] + \Psi \Psi^{\#} \rho D[n] - E[n] - \rho D[n]) - V(D[n])$$

$$= V(D[n] + \rho(\Psi \Psi^{\#} - I) D[n]) - V(D[n])$$

$$= ||D[n] + \rho(\Psi \Psi^{\#} - I) D[n]||^2 - ||D[n]||^2$$

$$= \text{Trace} \{ [D[n] + \rho(\Psi \Psi^{\#} - I) D[n]]^t \times [D[n] + \rho(\Psi \Psi^{\#} - I) D[n]] \} - ||D[n]||^2$$

$$= \text{Trace} \{ D[n]^t D[n] + \rho D[n]^t (\Psi \Psi^{\#} - I) D[n] + \rho D[n]^t (\Psi \Psi^{\#} - I) D[n] + \rho^2 D[n]^t (\Psi \Psi^{\#} - I)^t (\Psi \Psi^{\#} - I) D[n] \} - ||D[n]||^2$$

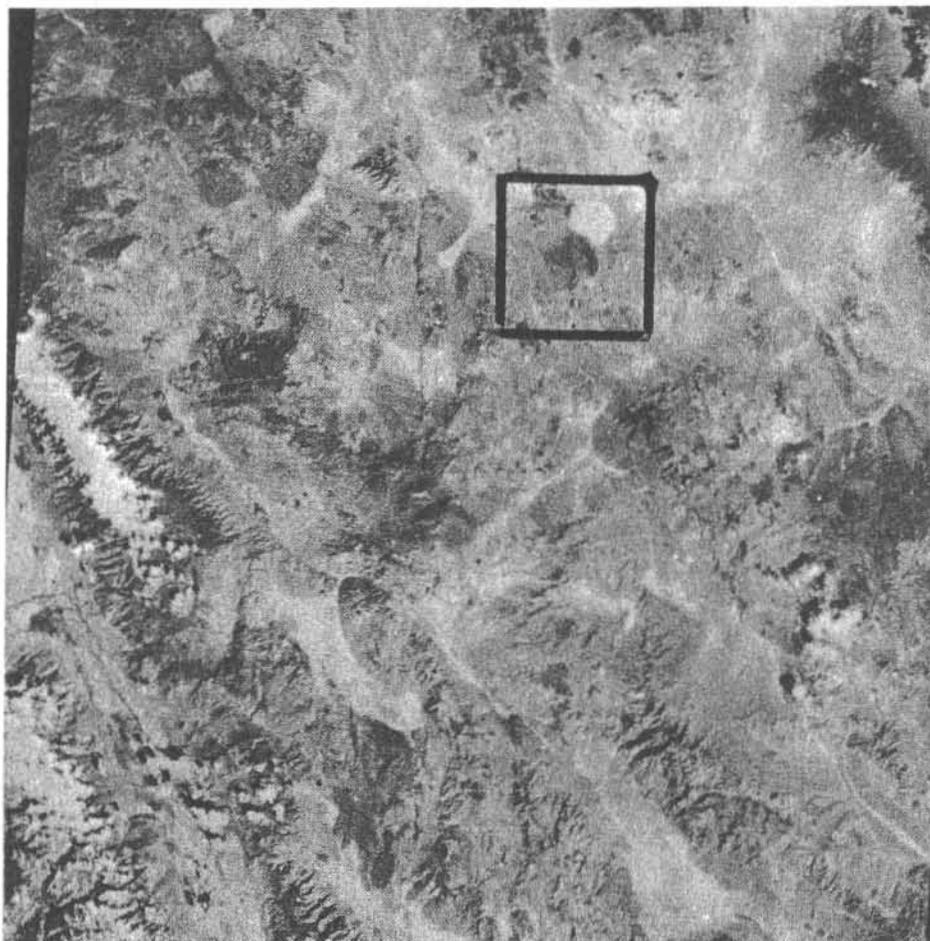
$$\begin{aligned}
&= ||D[n]||^2 + 2\rho \text{Trace} \{D[n]^t(\Psi\Psi^{\#}-I)D[n]\} \\
&\quad + \rho^2 ||D[n]||^2 - ||D[n]||^2 \text{ (by matrix identity b)} \\
&= 2\rho \text{Trace} \{D[n]^t(\Psi\Psi^{\#}-I)D[n]\} + \rho^2 ||D[n]||^2 \\
&= 2\rho \text{Trace} \{D[n]^t\Psi\Psi^{\#}D[n]\} - 2\rho \text{Trace} \{D[n]^tD[n]\} \\
&\quad + \rho^2 ||D[n]||^2 \\
&= -2\rho ||D[n]||^2 + \rho^2 ||D[n]||^2 \text{ (by matrix identity a)} \\
&= -||D[n]||^2(2\rho - \rho^2).
\end{aligned}$$

For  $0 < \rho \leq 2$ ,  $2\rho - \rho^2 = \rho(2-\rho) \geq 0$ , and  $\Delta V(D[n]) \leq 0$  for all  $D[n]$ . Also,  $\Delta V(D[n]) = 0$  if  $D[n] = 0$ .

By Lyapunov's stability theorem for discrete systems [6],  $\lim_{n \rightarrow \infty} V(D[n]) = \lim_{n \rightarrow \infty} ||D[n]||^2 = 0$

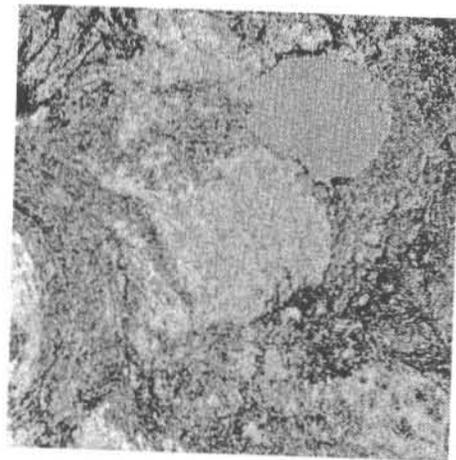
#### REFERENCES

1. J.D. Patterson and B.F. Womack, "An adaptive pattern classification system," *IEEE Trans. Systems Science and Cybernetics* SSC-2, 1 (1966).
2. A.Y. Hung, "An Introduction to Multiclass Pattern Recognition in Unstructured Situations," Div. of Eng. Research, MSU, Report No. 12, December, 1970.
3. J.S. Koford and G.F. Groner, "The use of an adaptive threshold element to design a linear optimal pattern classifier," *IEEE Trans. Inform. Theory* 12, 1 (1966).
4. J.D. Patterson, T.J. Wagner and B.F. Womack, "A mean-square performance criterion for adaptive pattern classification systems," *IEEE Trans. Automat. Control* AC-12, 2, 195-197 (1967).
5. W.G. Wee, "Generalized inverse approach to adaptive multiclass pattern classification," *IEEE Trans. Computers* C-17, 12 (1968).
6. R.E. Kalman and J.E. Bartram, "Control system analysis and design via the 'Second method' of Lyapunov, II Discrete-time system," *Trans. ASME J. Basic Engr.* 82, D, 2, 371-400 (1960).
7. R.H. Caron, "Evaluation of Full-Scene Registered ERTS MSS Imagery using a Multitemporal/Multispectral Bayes Supervised Classifier," Fourth Annual Remote Sensing of Earth Resources Conference, Uni. of Tennessee Space Institute, Tullahoma, Tenn., March, 1975.
8. W. Eppler, "Canonical Analysis for Increased Classification Speed and Channel Selection," Symposium on Machine Processing of Remotely Sensed Data, Purdue University, W. Lafayette, Indiana, June 3-5, 1975.
9. Clay Jones, "Implementation of an Advanced Table Look-up Classifier for Large Area Land-use Classification." Ninth Symp. on remote sensing of the Environment, Ann Arbor, Michigan April 1974.
10. H.P. Friedman and J. Rubin, "On some invariant criteria for grouping data," *J. Amer. Statist. Assoc.* 62, 320, 1159-1178 (1967).
11. G.H. Ball and D.J. Hall, "ISODATA, a novel method of data analysis and pattern recognition," Technical Report, Stanford Research Institute, Menlo Park, California (1965).
12. Lawrence C. Rowan, et al., "Discrimination of Rock Types and Altered Areas in Nevada by the use of ERTS Images", Geological Survey Professional paper 883, 1974.
13. A.F.H. Geotz, et al., "Application of ERTS Images and Image Processing to Regional Geologic problems and Geologic Mapping in Northern Arizona," Tech. Report 32-1597, JPL, May 15, 1975.
14. Roger P. Ashley, "Goldfield Mining District", Report 19, Nevada Bureau of Mines and Geology, March, 1974.
15. N.J. Nilsson, "Learning Machines", McGraw-Hill, New York, 1965.



1072-18001- 5 TRW/SPF 23MAR75

Figure 2. TRW System Corrected Goldfield Scene - The Area Inside The Box is Extracted for Classification.



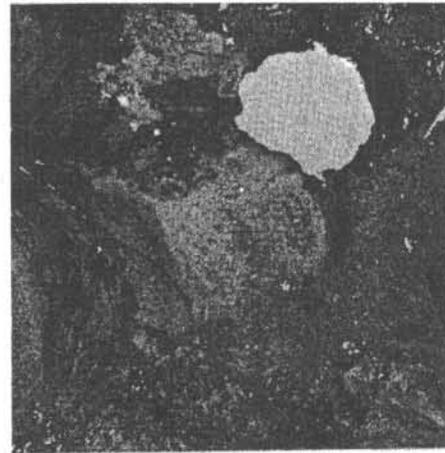
GOLDFIELD BAND 51072-18001

DENSITY SLICE TM TRW/SPF

Figure 3. Goldfield Nevada Subscene and Density-slicing Thematic Map



LSE CLASSIFICATION MAP



BAYES CLASSIFICATION MAP TRW/SPF

Figure 4. LSE and Bayes Classification Map.