Theses   Thesis/Dissertation Collections

5-20-1989

Selection and analysis of optimal textural features for accurate classification of monochrome digitized image data

Denis J. Robert

Follow this and additional works at: http://scholarworks.rit.edu/theses

SELECTION AND ANALYSIS OF OPTIMAL

TEXTURAL FEATURES FOR ACCURATE CLASSIFICATION OF

MONOCHROME DIGITIZED IMAGE DATA

by
Captain Denis J. Robert
B. Eng / M. S. Candidate

Canadian Armed Forces

Rochester Institute of Technology
Center for Imaging Science

20 May 1989

A thesis submitted in partial fulfillment
of the requirements for the degree of Master of Science
in the Center for Imaging Science in
the College of Graphic Arts and Photography of the
Rochester Institute of Technology

Signature of the Author   Captain Denis J. Robert
                          Center for Imaging Science

Approved by _____ Name Illegible   6-5-89
                       Coordinator, M. S. Degree Program

i

CENTER FOR IMAGING SCIENCE
COLLEGE OF GRAPHIC ARTS AND PHOTOGRAPHY
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

<u>CERTIFICATE OF APPROVAL</u>

M.S DEGREE THESIS

The M.S. Degree Thesis of Captain Denis J. Robert
has been examined and approved by the thesis committee
as satisfactory for the thesis requirement for the
Master of Science Degree

Dr. John Schott, Thesis Advisor

Dr. Roger Easton

Mr. Carl Salvaggio

/ June 89
Date

ii

THESIS RELEASE PERMISSION FORM

ROCHESTER INSTITUTE OF TECHNOLOGY
COLLEGE OF GRAPHIC ARTS AND PHOTOGRAPHY
ROCHESTER, NEW YORK

I, Denis J. Robert, prefer to be contacted each time a request for reproduction is made. I can be reached at the following address:

> Denis J. Robert
> 260, Rang 7 E.
> Thurso, Quebec
> JOX 3BO
> Canada

Date _____ 1 June 89 _____

SELECTION AND ANALYSIS OF OPTIMAL TEXTURAL FEATURES FOR ACCURATE CLASSIFICATION OF MONOCHROME DIGITIZED IMAGE DATA

by

Denis Robert

Submitted to

the Center of Imaging Science

in partial fulfillment of the requirements for

the Master of Science Degree

at the Rochester Institute of Technology

Abstract

A feature-selection technique based on measures of global class separability in multidimensional feature space is proposed for classifying monochrome digitized imagery by machine. Feature-selection procedures are an essential step in optimal classification in reduced feature space. Textural features constitute the type of measurements used to characterize image data due to its monochrome nature.

The ability of the proposed feature-selection technique to provide an optimal environment for classifying image pixels is measured by the Gaussian Maximum Likelihood method. The appropriateness of using textural features to characterize monochrome digital image data is assessed in similar fashion. The robustness of the proposed feature selection technique, and that of use of textural features, to provide for accurate and effective image processing is tested by analyzing several monochromatic images which contain multiple ground-cover classes, various resolutions, orientations, grey-level quantization levels, and individual textural feature parameter settings.

Acknowledgments

I wish to thank all the members of the Digital Imaging and Remote Sensing Lab for their assistance and support throughout the duration of this study. In particular I wish to acknowledge :

Dr. John Schott, whose guidance, patience, and concern were invaluable to the success of this project,

Dr. Roger Easton, who patiently corrected all the Canadian grammer used throughout this document,

Carl Salvaggio, whose timely counselling and friendly disposition were always greatly appreciated,

Eugene Kraus, whose assistance in developing and implementing pertinent software contributed significantly in optimizing code processing efficiency and accuracy,

Steve Shultz, whose patience was appreciated when my numerous and never-ending batch jobs constantly occupied the resources of "his computer".

DEDICATION

This thesis is dedicated to

all my close friends

within the Imaging Science department at

the Rochester Institute of Technology

who made the experience in Rochester one I shall always

cherish and remember.

I wish you all the very best throughout your studies,

and in your future careers.

TABLE OF CONTENTS

xi

LIST OF TABLES

LIST OF FIGURES

Figure 4.4-19   Global Classification Accuracy of Test Data
for Analysis 31 for Subsets of Features

1.0 Introduction

For many years, remotely sensed data has allowed man to examine the state of this planet's resources and to unlock some of the earth's more puzzling mysteries. Observational instruments carried by aerial and spaceborne platforms generate immense quantities of image and numerical data which permit man to better understand and manage his environment.

With the advent of the digital computer in the early 1950's, the science of remote sensing merged the pictorial and quantitative aspects of image information. As the value of remotely sensed data became more dependant upon timely processing, and while technology made available increasing amounts of information, the importance of numerical systems progressed rapidly. Today's digital computers are indispensable tools to achieve the processing speed, economy, precision and efficiency required to shape information in accordance with the requirements of man.

The computer is of particular value in remote sensing for processing digitally encoded images to replace time-consuming and laborious visual analysis. Such processing consists of automated manipulation and interpretation of image data, including image rectification, restoration, enhancement, and classification. This study shall make extensive use of computer processing to classify

1

digital image data, and to select image features to optimize the classification process.

We first examine the concepts which pertain to image pattern classification and to optimal feature selection.

1.1   Digitally Encoded Images

The signals collected by a sensor due to the electromagnetic energy reflected by and emanating from the earth's surface are processed and recorded in pictorial format to form digital images. Today's digital images are of high quality and are stored in a computer usually as two-dimensional arrays, where each picture element or pixel has an associated grey-tone value. Digital images may vary in size and are usually square to simplify processing. Figure 1.1-1 depicts a 5x5 digital image that possesses two grey-tone values, 0 and 1 (0 corresponding to light, 1 corresponding to dark), where the arrangement of pixel brightness counts represents the number four.
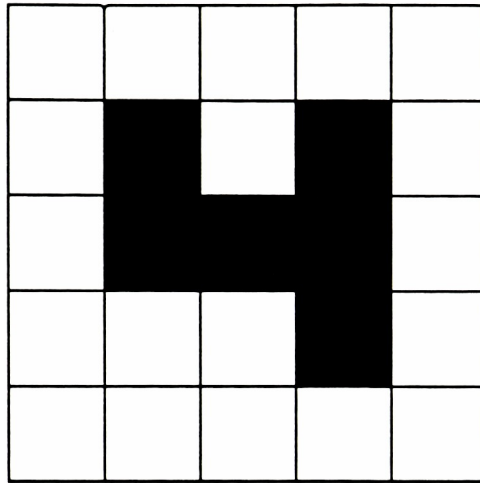
Figure   1.1-1

To put an image in a form suitable for computer processing, it must be digitized (i.e. represented as integral spatial and brightness coordinates). Integer spatial coordinates are assigned to equally spaced samples of the image, and each pixel is ascribed a brightness value from a limited number of discrete grey levels. Gonzalez and Wintz (1987) give an excellent description of digital images and of the digitizing process.[1]

1.2   Image Pattern Classification

Digital image classification consists of collecting and analyzing digital image data, and applying statistically and/or structurally

based decision rules to determine the identity of its components. The decision rules can be based upon spectral radiances measured from imaged ground targets (spectral), upon the geometric shapes, sizes, and patterns present in the image data (spatial), or upon the temporal variations in spectral radiance (temporal). Figure 1.2-1 illustrates a model of an image pattern classification system, as given by Swain and Davis (1978).[2]
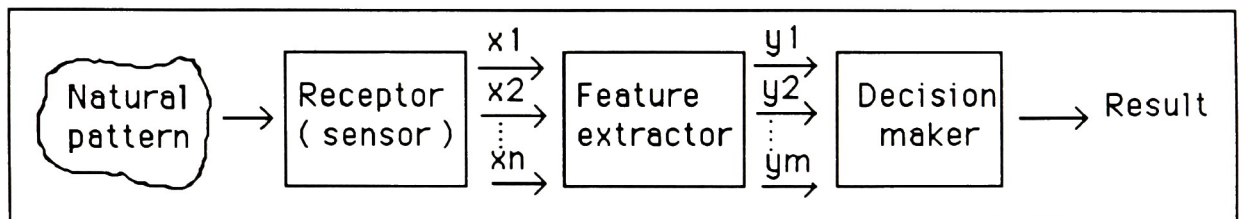
Figure 1.2-1

In Figure 1.2-1, the receptor may be an airborne or satellite-based sensor which detects electromagnetic energy from natural ground cover patterns. The output of the receptor is a set of " n " simultaneously collected measurements, corresponding to each channel of the sensor, from which " m " features are extracted to discriminate between image pixels. Features consist of original or derived measurements of radiance which are translated into an m-dimensional feature space. The classifier assigns vectors in this space to one of several classes based on a classification rule.

1.2.1 Statistical and Structural Classification

Both statistical and structural approaches exist for classifying digital image data. In statistical classification, a set of characteristic features is extracted from each image pixel or image block. Image pixels are attributed to separate image classes, usually by partitioning the associated multidimensional feature space, as reported by Fu (1968).[3]

The structural approach consists of describing an image pattern using a hierarchical system of structures and grammar rules. These latter entities are broken down into sets of subpatterns, as suggested by Pavlidis (1977), which are described in their simplest form by other subpatterns called primitives.[4] Syntactic placement rules describe the allowed structural relationships of image primitives and subpatterns in each image pattern, and are the basis for classification. The attractiveness of the structural approach for classifying depends on the construction and description of the primitives. In many recognition problems involving complex patterns, the number of features required to ensure accurate classification is sometimes very large and a structural approach is judged more practical.

The structural approach shall not be used in this study, as use of a grammar hinders its effective implementation. In addition, the local

order or repetitive pattern is rarely constant from place to place and phase shifts, twisting and stretching are distortions that locally affect spatial periodicity. It is difficult to identify pattern primitives in even perfectly repetitive image pattern since the primitives may themselves consist of other subpatterns.

A novel approach to describing and modelling image structures uses mathematical curves known as fractals to characterize image objects with a hierarchical organization of scaled structures. Fractals were developed by Mandelbrot (1980) which enabled him to organize a whole universe of self-similar objects in a mathematical fashion.[72] Further research in this field of study has resulted in development of fractal numbers which measure local image textures. Image fractals shall not be employed in this study as they have not yet to been widely used for accurate and effective classification of digital image data.

While the statistical approach to pattern recognition is not perfect, it is particularly appropriate for classifying remotely sensed image data into ground cover classes, as reported by Swain and Davis (1978).[5] Statistical decision theory accounts for inherent variations in remotely sensed data, and attempts to reduce their negative effects on classification accuracy. This approach is also very tolerant of errors associated with the questionable identity of available training samples, and provides for image pattern

classification which is most probably correct when pattern classes actually overlap in feature space (i.e. some measurements in feature space are indistinguishable between classes). The majority of today's commercial recognition systems use statistical approaches, as reported by Devijver and Kittler (1986), because often, only statistical tools can allow man to comprehend the extreme variability of image patterns, due to the randomness of nature.[6]

Of interest to the reader is the introduction by Fu (1986) of novel grammars which unify the statistical and the syntactic approaches to digital image pattern classification.[7]

1.3   Statistical Image Pattern Classification - Strategies

Statistical image pattern classification is a quantitative and automated decision-making process. Measurements collected by a sensor are assumed to have a density function or statistical distribution in feature space, conditioned on the pattern class. Pattern feature vectors are viewed as observations drawn from respective class-conditional density functions.

Figure 1.3-1 shows various strategies to design a statistical classifier, depending on the nature and on the reliability of

information concerning image class-conditional density functions. Selection of the appropriate strategy must ensure maximum classification accuracy of all image data and efficiency of the associated process. Different classification strategies, employed over a common image area, will most often result in different degrees of classification accuracy and efficiency.
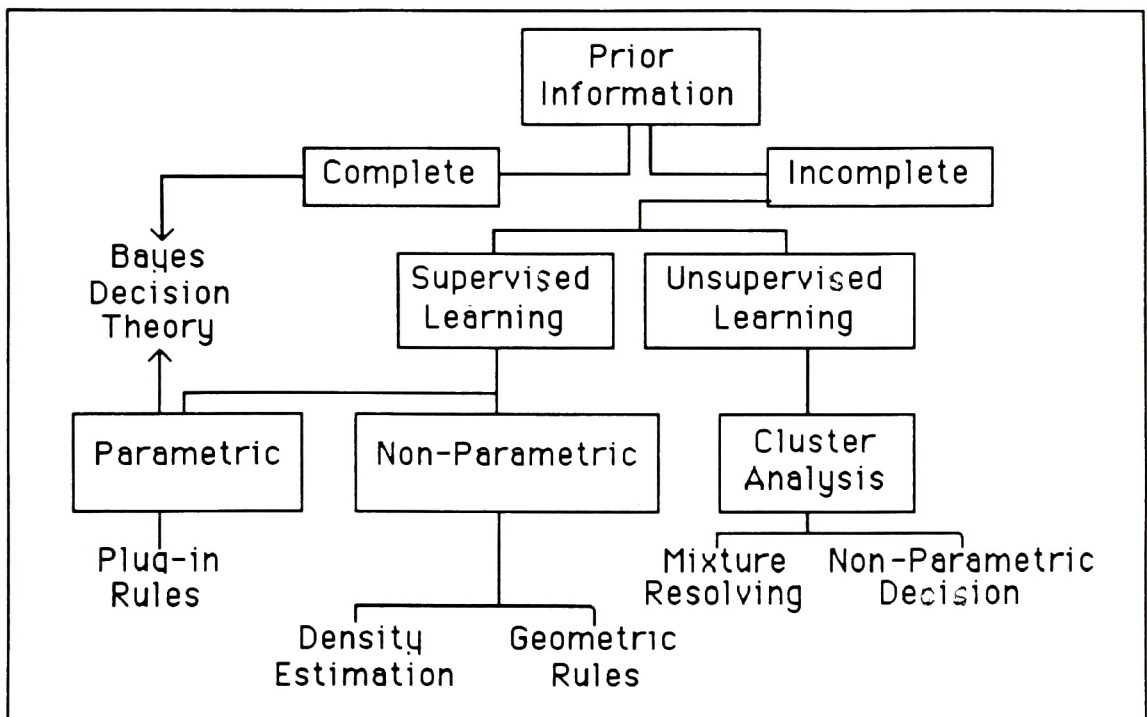
Figure   1.3-1

As the information available concerning image class probability distributions is often incomplete, two strategies exist to extract quantitative information from image data and establish the nature of class density distributions: the supervised method, and the unsupervised method.

1.3.1   Supervised Classification

The basic assumption in statistical supervised classification is that all ground cover classes can be described by a probability distribution in feature space.   The analyst supervises the establishment of decision boundaries within the classification environment by providing training samples to the classifier.   This type of classification is described as a five-stage process by Richards (1986) and is reviewed below.[8]

In the first stage, ground cover classes are identified based on the available features.   Ground cover classes include urban, croplands, rangelands, etc.   Extensive visual analysis of the imagery is required prior to attributing class names to ensure maximum accuracy and efficiency of the classification process.

In the second stage, training pixels are selected from homogeneous and representative areas in the image for each class; where the full

range of variability in grey tone values for each class is included. If homogeneous areas are not readily available, the analyst must select more than one region for a given class or must apply thresholding algorithms to the data of interest to remove unwanted image patterns, as stated by Maxwell (1976).[9] Figure 1.3.1-1 illustrates possible image sampling areas for supervised classifiers.
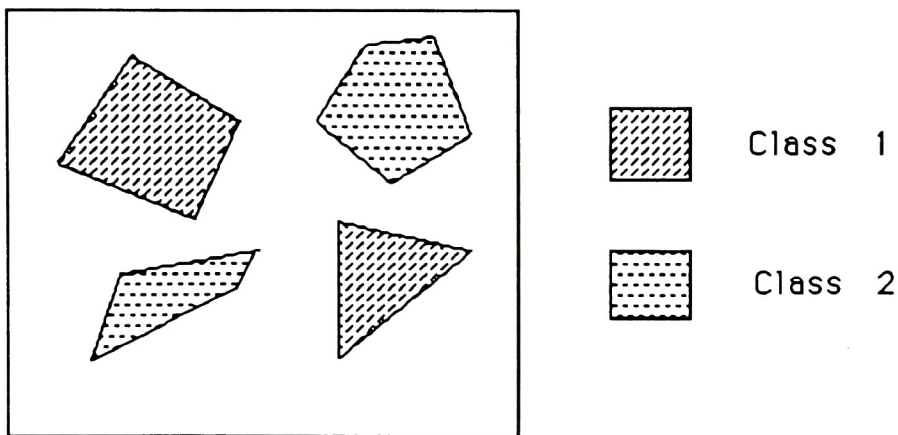
Figure 1.3.1-1

Classifier training is the third stage; class parameters are estimated from the chosen training sets. These parameters consist of the statistical properties of each class distribution in feature space (the parametric approach), or serve to define equations which partition the feature space about each class (the non-parametric approach). Assuming the class probability distributions are

multivariate normal, these parameters consist of class-mean vectors and class covariance matrices. These statistics describe the average position and the amount of spread of class density functions about their mean vector in feature space.

In the fourth stage, every image pattern is labelled into one of the predetermined classes using the trained classifier. As the number of image pixels to classify is often overwhelming, the computer becomes an essential tool to achieve process efficiency.

The fifth and final stage consists of producing tables of data (i.e. confusion matrices) and plots which portray the results of the classification process.

1.3.2   Unsupervised Classification

Unsupervised classification differs in that classes of interest are no longer specified based on available features, but are defined by their separability in feature space (i.e. image classes are identified in feature space following delineation of inherent data structures). Unsupervised classification produces reliable results when image ground cover classes are easily discriminated in feature space, and is used when reliable training data for supervised classification is impossible or expensive to obtain.

Training data for the unsupervised process is chosen to include a heterogeneous mixture of all classes and accounts for their within-class variabilities. Figure 1.3.2-1 illustrates possible sampling areas for training unsupervised classifiers.
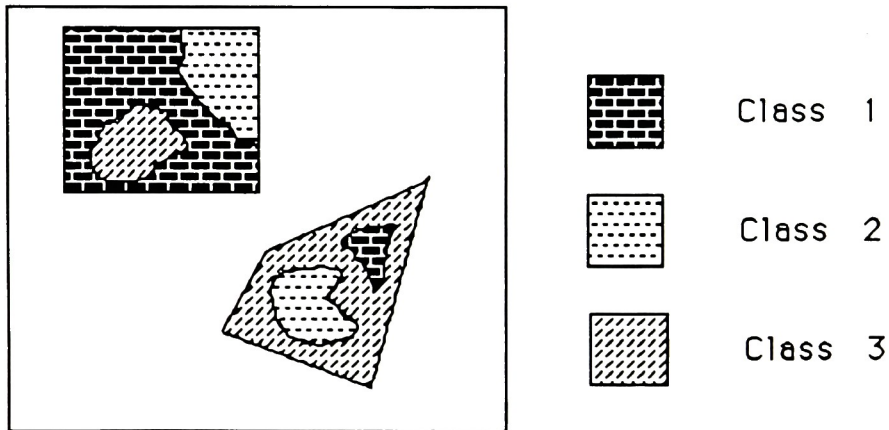
Figure   1.3.2-1

The image training data are submitted to automated clustering processes in feature space which determine natural groupings of classes, as reported by Richards (1986).[10] Each cluster contains data which is described by the probability distribution of at least one class. Clusters which contain more than one class are usually separated at the discretion of the analyst using tools such as maps, air photographs, or site visits.

A wide choice of clustering algorithms are available to the analyst. Methods such as CLASS by Fromm (1976)[11] and ISODATA by Kan (1972)[12] are reviewed by Bryant (1979).[13] One of the more common methods was described by Duda and Hart (1973) and is known as the K-means algorithm.[14] In this process, an initial set of decision boundaries is established in the feature space. Each class is assigned a mean vector and each training pixel is assigned to the closest of these. A new set of mean vectors is then calculated from the initial data and image pixels are then reassigned. This process continues until each mean vector and associated class cluster remains relatively stationary in feature space. A statistical classifier can then be applied to the entire image and class parameters can be estimated from each individual cluster, as reported by Schowengerdt (1983).[77]

In both supervised and unsupervised classification, the task of the classifier is to establish a set of discriminant functions which divide the image feature space into appropriate decision regions. Discriminant functions are defined such that the discriminant function for the nth pattern class has the largest value at every point in that part of feature space.

Supervised and unsupervised classification algorithms are used extensively with image classifiers. These classifiers can be divided into two groups; non-parametric and parametric.

1.3.3 The Non-Parametric Classifier

The non-parametric classifier makes no distributional assumptions with respect to the data. The analyst must either estimate the class density functions or use some non-statistical decision rule to partition feature space, as stated by Devijver and Kittler (1987).[15] Non-parametric classifiers are sometimes used in remote sensing exercises, but are encountered much less frequently than their parametric counterparts. Non-parametric classifiers are more powerful in their ability to estimate probability functions but their computational costs are often excessive and require large numbers of training patterns, as reported by Swain and Davis (1978).[17] These limitations often prove to be unacceptable for many classification exercises. Schowengerdt (1983) states that non-parametric minimum distance and parallelepiped algorithms used to establish non-statistical decision rules in feature space, generally lack classification accuracy despite being computationally efficient and easy to implement.[73] These classifiers are preferred when the nature of the data's distribution is unknown.

Non-parametric classifiers are neither considered nor reviewed in this study as all class density functions are assumed to possess gaussian distribution in feature space. The following sections provide argument for this assumption. For a detailed description of

several non-parametric classifiers, the interested reader is directed to Nilsson (1965).[16]

1.3.4 The Parametric Classifier

Parametric classifiers assume that the form of class probability distributions is known, but that some of the distribution parameters are undefined, as reported by Devijver and Kittler (1986).[18] The unknown parameters are generally replaced by values estimated from representative training data sets prior to image pattern classification. The class probability distributions are generally assumed to be gaussian in nature, which often results in efficient and accurate classification of image target data from various ground cover classes.

1.3.5 Preferred Approach for Image Pattern Classification

The parametric supervised method of image pattern classification (supervised partitioning of feature space and parametric decision rules) shall be the preferred approach in this study . A multivariate gaussian or normal model is assumed for the distribution of points for each class and each image pattern is assigned to the class in which it has the greatest probability of belonging.

1.4   Class Probability Distributions / Normality
        Assumptions in Feature Space

The assumption of multivariate gaussian probability distributions for classes in feature space is particularly appropriate for remotely sensed data of earth features, as reported by Swain and Davis (1978).[19]   Although this rationale imposes a restriction on the nature of class density functions, it is known that most remotely sensed natural processes can be modelled by a multivariate normal behaviour, or can be described by a combination of gaussian processes.   Even when this assumption is severely violated, classifiers designed on this basis are often still very correct. Fukunaga and Flick (1986) state the difficulties of verifying the gaussian nature of multivariate data when assuming normal class distributions in multidimensional measurement space.[20]   Johnson and Wichern (1982) report that one and two-dimensional investigations are ordinarily sufficient to assume normality in multidimensional feature space.[76]

The Bayes Optimal or Gaussian Maximum Likelihood classifier is a parametric classifier which uses the assumption of normality of the statistical properties of image classes in feature space to minimize the mean error of incorrect classification.

1.5   Image Training and Test Data

The image data which characterizes each ground cover class in feature space is commonly divided into two subsets: image training data and image test data . The image training data is used to design the classifier while the image test data is set aside to assess classification accuracy.

Training samples are used to determine the values of class distribution parameters, such as class mean vectors and class covariance matrices, and serve to establish decision boundaries in feature space prior to execution of the classification process. Training samples represent particular class feature spaces and are used to build the classifier.

Schowengerdt (1983)[21] and Harris (1987)[22] stress the importance of properly selecting training samples for supervised classification of image patterns for credible results. Points of concern include attention to homogeneous training sampling areas, random selection of sample pixels, adequate numbers of training samples per class, whether the sample data possesses multivariate normal distribution, and the range of variability of training samples in each class. When the above requirements are not met, more than one training region is often required per image class. The theoretical minimum number of training samples is n+1 per class for n features;

10n or 100n training samples per class lead to better estimates of parameters, as reported by Swain and Davis (1978).[23]

As in the case of the image training data, the image test data must be representative of the data to be classified. The test data should ideally be chosen at random from the image data and include samples of the total areas under scrutiny. The larger the set of test data, the more representative the testing of the classifier. However, the cost of collecting these data and ensuring that they represent desired classes often restrict availability. Therefore, the analyst must select all image training and test data with care to minimize the costs and maximize the efficiency of the classification exercise without sacrificing classification accuracy.

1.6   Bayes Optimal / Maximum Likelihood Classifier

The Bayes Optimal or Maximum Likelihood Classifier is a parametric classifier which offers optimal performance over an entire classified data set if all image classes have unimodal, (single maximum) gaussian distributed probability density functions, as reported by Duda and Hart (1973).[24] As multi-modal class distributions generally cannot be adequately approximated by a single unimodal normal density function, optimal performance of the Bayes Optimal classifier is denied with exception to subdividing

multi-modal classes into subclasses, one for each mode of the actual distribution. The probability distribution for each subclass is represented by a normal density function in feature space. This classifier is preferred for use in this study due to its parametric nature, and to its particular suitability for normally distributed image data.

The basic strategy of the Bayes Optimal classifier is to minimize the *expected average loss* over an entire set of classifications to be performed as reported by Richards (1986).[25] The *expected average loss*, $L_X(W_i)$, is a measure of the penalty incurred when the classifier erroneously labels an image observation vector X as belonging to a class W_i when in reality it belongs to class W_j. To classify X into a class W_i, for i = 1, 2, 3, ..., m, using the Bayes Optimal classifier, $L_X(W_i)$ computed for each class W_i is derived as per equation (1) :

$$L_X(W_i) = \sum_{j=1}^{m} Cost(i,j) * p(W_j/X), \qquad (1)$$

where :

    Cost(i,j) : cost of classifying an image pattern with feature
                 vector X into class W_i when it is from class W_j, and

$p(W_j/X)$: probability that an image pattern is a member of class W_j given that it has a feature vector X in multidimensional feature space.

The cost function Cost(i,j) is specified by the analyst to suit the classification process at hand. A well known 0-1 cost function is most commonly employed where Cost(i,j) is equal to 0, or no cost, for correct classification and is equal to 1, or unit cost, for incorrect classification. Using this function, the *expected average loss* $L_X(W_i)$ is derived as per equation (2) :

$$L_X(W_i) = 1 - p(W_i/X), \qquad (2)$$

and will be minimized if W_i is selected so that $p(W_i/X)$ is maximized. Because of this maximizing property this classification strategy is also known as the Maximum Likelihood Classifier. Nilsson (1965) gives an excellent account of the derivation of the Maximum Likelihood decision rule based upon the 0-1 penalty functions.[26]

The Bayes Optimal or Maximum Likelihood classifier relies upon *a posteriori* probabilities $p(W_i/X)$ to classify image patterns. The *a posteriori* probability that an image pattern belongs to class W_i given that it has feature vector X is given by :

$$p(W_i/X) = \frac{p(X/W_i) * p(W_i)}{p(X)}, \qquad (3)$$

where :

        $p(X)$: summation of $p(X/W_i)p(W_i)$ over all classes.

As $p(X)$ is the same for all classes it can be ignored from further calculations. The probabilities $p(X/W_i)$ are known as *state-conditional* probability density functions and represent the probability that an image pattern has a feature vector X given that it is in class W_i. The probabilities $p(W_i)$ are known as *a priori* probabilities and represent the probability that class W_i occurs in the image area of interest. The product of the *state-conditional* probability $p(X/W_i)$ and the *a priori* probability $p(W_i)$ represent the probability that an image pattern has a feature vector X and is in class W_i.

The *a priori* probabilities can be estimated from outside sources of information such as ground surveys, existing maps or historical data. In practice, these probabilities are often difficult (if not impossible) to obtain. Therefore, they are often assumed to be equal for all classes and can consequently be removed from further calculation. Using the 0-1 loss function and assuming all above assumptions are valid, the Bayes Optimal or Maximum Likelihood

classification rule assigns the feature vector X to class W_i when the conditions of equation (4) are satisfied .

$$D_{W_i}(X) = p(X/W_i)_{max}, \qquad (4)$$

where :

$$p(X/W_i)_{max} = \text{maximum value of all } p(X/W_i)$$

$$\text{over all } i = 1, 2, \ldots\ldots, m.$$

The feature vector X is therefore assigned to the class W_i that produces the maximum value of $D_{W_i}(X)$. $D_{W_i}(X)$ is known as a *discriminant function* which partitions feature space into separate classes for classification of image data.

Assuming that all *state-conditional* probability distributions are multivariate gaussian leads to a Bayes Optimal classifier which is computationally efficient and relatively easy to implement. In addition, this classifier requires that relatively few parameters need be determined for each class, thus reducing the requirement for large numbers of training samples. *State-conditional* probability distributions, in a K-dimensional normal feature space, are expressed in general multivariate form by equation (5) :

$$p(X/W_i) = \frac{1}{[(2\pi)^{K/2}] * [/E_i/^{0.5}]} * \exp[-0.5(X-M_i)^{t*}E_i^{-1*}(X-M_i)], \qquad (5)$$

where :

E_i : $K*K$ symmetric covariance matrix for class W_i,

$/E_i/$: determinant of E_i,

$E_i{}^t$: transpose of E_i,

$E_i{}^{-1}$: inverse of E_i,

X : K-dimensional image pattern feature vector, and

M_i : K-dimensional mean vector for class W_i.

Each class covariance matrix E_i and class mean vector M_i is estimated from training samples. In this study, the Moore-Penrose generalized inverse of E_i was derived when $E_i{}^{-1}$ did not exist (singular matrix).[75] This method of computing the inverse of a singular matrix consists of providing a solution from a slightly modified non-singular version of E_i. Appendix D refers.

Assuming that all *a priori* probabilities are equal, the discriminant functions of equation (4) are derived by taking the natural logarithm of each side of equation (5) to produce :

$$D_{W_i}(X) = A_i - 0.5*[(X-M_i)^t * E_i{}^{-1} * (X-M_i)], \qquad (6)$$

where

$$A_i = [-0.5 * \ln/E_i/] - [K/2 * \ln(2\pi)]. \qquad (7)$$

This result was given by Duda and Hart (1973).[27] To make the decision rule of equation (6) computationally simpler, the analyst

may assume that the covariance matrices E_i for each class W_i ($i = 1$, 2, ..., m) are equal. When such assertion is valid, a pooled covariance matrix E_p is derived for all classes :

$$E_p = \frac{[E_1{}^*n_1 + E_2{}^*n_2 + ... + E_m{}^*n_m]}{[n_1 + n_2 + ... n_m - m]} \qquad (8)$$

where :

n_i : number of training samples in each of the

$i = 1, 2, ..., m$ classes.

The term A_i in equation (7) is then a constant for all classes so that the discriminant function $D_{Wi}(X)$ in equation (6) becomes :

$$D_{Wi}(X) = -0.5^*[(X-M_i)^t {}^* E_p{}^{-1} {}^* (X-M_i)], \qquad (9)$$

and the feature vector X is assigned to the class that results in minimum discriminant function. This discriminant function is known as the *Mahalanobis distance*. Pooled covariance matrices offer increased classification efficiency compared to individual class covariance matrices, but also result in decreased classification accuracy when they are used. Both approaches are tested in this study; however, use of individual covariance matrices is preferred to maximize classification accuracy.

24

Derivations and treatments of the Bayes Optimal classifier are given by Schowengerdt (1983)[28] and Andrews (1972).[29]

1.7   Image Preprocessing

Preprocessing of digital images prior to pattern classification is an essential operation which can greatly increase the accuracy of the classifier.   Image preprocessing consists of geometric and radiometric image corrections, noise removal, and image enhancement.   These operations correct changes in image shape, radiometric distortions due to a changing atmosphere, unwanted signals and enhance visual interpretability by increasing the contrast between image patterns.   The extent of image preprocessing on an image is dictated by the sensor's characteristics.   Several imaging preprocessing techniques are used in this study and deserve our attention.

Geometric correction provides for compensation of image distortions introduced by variations in the altitude, attitude, and velocity of the sensor platform, the earth's curvature, atmospheric refraction, relief displacement, and scan nonlinearities.   Predictable distortions are corrected by applying formulas derived from mathematical modelling of the sources of distortion using orbital models and scanner calibration data, as reported by Anuta (1973).[30]

Unpredictable distortions are corrected by identifying ground control points occurring in the image (pixel rows and columns) and measured from a map (latitudes and longitudes), which are used to determine coordinate transformation equations. These equations relate geometrically correct map coordinates to distorted image coordinates and provide for transformed image geometric integrity.

Radiometric correction preserves maximum image resolution and provides for correction of radiometric distortions caused by scene illumination, atmospheric conditions, viewing geometry, and instrument response characteristics. Techniques for radiometric correction include contrast manipulation operations on image grey level histograms (i.e. linear contrast stretching, histogram equalization and level slicing, cyclic contrast enhancement, and thresholding). These operations consist of either individual pixel-by-pixel radiometric transformations or of adaptive algorithms whose parameters change from pixel to pixel in an image according to local contrast, as reported by Fahnestock and Schowengerdt (1983).[31] Of particular interest to this study are histogram equalization or flattening, grey-scale thresholding, and histogram level slicing preprocessing operations.

In histogram equalization, the probability of occurrence of each image grey level is made equal; this removes the effects of unequal overall contrast and brightness in the initial scene. Because most

image histograms are gaussian in shape, this operation tends to reduce contrast in the very light or dark image areas, and expands the middle grey levels toward the low and high radiances. The resulting image histogram is approximately uniform, as reported by Gonzalez and Wintz (1977).[32] In unprocessed images, such effects could otherwise dominate the measured feature values. Weszka, Dyer and Rosenfeld (1976) discuss several histogram flattening algorithms available to the analyst.[33]

Image grey scale thresholding allows the analyst to recognize and manipulate individual objects from a given image by ensuring that each of the latter entities possesses discrete boundaries, as reported by Troy, Deutsch and Rosenfeld (1972).[34] The resulting image is segmented into classes defined by a single grey level threshold. This procedure is most effective when applied to those images where the range of grey tones is confined to a given range, and when object and background grey tones do not overlap. When grey-value range overlap does exist, object isolation is no longer assured.

Histogram level slicing is an enhancement technique that divides the scale of image grey levels into individual groups and attributes single grey tones to each. The number of output grey tones is established by the analyst. When properly done, level slicing can reduce unwanted signals and enhance textural information.

Lillesand and Kiefer (1987) describe other image preprocessing algorithms to eliminate undesirable image characteristics.[35] All image preprocessing operations should introduce no biases into feature measurements and increase the accuracy of the classifier.

1.8   Evaluation of Classifier and Optimal Feature Space

The statistical approach to pattern recognition and the Bayes Optimal classifier were selected for this study for several reasons. The main reason is that class overlap in feature space exists. When classification errors occur (especially when classes overlap), the Bayes Optimal classifier minimizes the associated probability of error.

The probability of error is a valuable indicator of the degree of confidence the analyst can have in classification results, as reported by Swain and Davis (1978).[36] Should this error be unacceptably high, the analyst may seek other features to discriminate among the classes of interest, or change the design of the Bayes Optimal classifier to maximize classification accuracy. Swain and Davis (1978) propose various measures of the probability of classifier error.[37] Measures of the minimum overlap area between class density functions were difficult to implement in feature spaces with dimension greater than one. Error estimation

based on classification accuracy of test and training data required sufficient representative data points to be acceptably accurate. Costs of acquiring image data also deserved consideration. Although classifier error derived from training data is an overly optimistic estimator of classification accuracy, test data allows for derivation of a more reliable although still highly biased measure.

To reduce misclassification, the analyst may choose to apply probability thresholds to class probability distributions in feature space.[74] Such approach does not improve classification accuracy of image pixels within class boundaries; but prevents misclassification of pixels outside the boundaries. Thresholds improve estimates of the total area covered by each class within a given image area.

Other reliable and accurate estimators of classifier error are derived from divergence statistics, which constitute measures of statistical separability between all classes in feature space. The greater the value of the latter statistics, the more effective the classifier (i.e. the lesser the probability of classification error). Divergence-based statistics such as the Jeffries-Matusita (J-M) or Bhattacharrya distance, proposed by Wacker (1971)[38], transformed divergence, proposed by Swain and Davis (1978)[39], and class separation metrics generated from weighted class divergence matrices, proposed by Schott et al. (1988)[69], measure class

separability in terms of image class means and covariances. The first two measures are bounded by the probability of misrecognition (or the Bayes risk), as shown by Chen (1973).[40] Transformed divergence is almost as effective as the J-M distance, is more economical, and considerably superior to simple divergence, as reported by Swain et al. (1971).[41] Class separation metrics are chosen in this study and are generated from the Bayes Optimal classification strategy which ensures minimum global classification error over an entire set of normally distributed image data in feature space.

These measures not only assess the accuracy of the classifier, but also serve to validate the selection of optimal features. By measuring the accuracy of the classifier, the ability of features to ensure optimal partitioning of feature space into classes is also ascertained. This shall be further discussed in section 2.5.1.

A final measure of classifier efficiency is the processing time required to classify image patterns. This statistic corresponds to the time needed to read in pertinent data, make classification calculations, and write results in legible form. Should processing time prove to be unacceptably long, the analyst may consider redesigning the classifier to reduce costs, at the expense of possibly sacrificing on the accuracy of the classifier.

The three methods preferred for evaluating the classifier are classification of randomly selected image test and training data, analysis of class separation metrics derived from weighted class divergence matrices, and computation of required processing times to perform the identification exercise. These approaches were selected from other proposed methods due to their effective implementation, their availability from the developed software, and their accurate and reliable nature.

2.0 Digital Image Features

Computer classification of remotely sensed image data is a quantitative process where image pixels are identified based upon their feature values. In statistical classification, proper characterization of image contents depends on the analyst's ability to determine the features to be measured. A feature is a measurement, or a mathematical transformation of measurements, made from an image, which serves to classify image components into appropriate classes. Features must properly describe the contents of an image, remain unaffected by image translation and rotation, and provide for easy processing. The analyst usually decides on the nature and number of features to be measured from the physics and complexity of the imagery.

To summarize previous discussion, a feature vector X of an image pixel or image pattern is represented by an n-dimensional measure in feature space. Each dimension corresponds to a specific spectral, temporal or spatial measurement extracted from an imaged target, or a mathematical transformation of these entities, as stated by Schowengerdt (1983).[42] An example is an image pattern characterized by a set of grey levels in a multispectral image space. Each spectral band corresponds to a specific dimension of the image pattern feature vector. The dimensionality of the feature vector is chosen at the discretion of the analyst. Figure 2.0-1 shows a three-

dimensional feature space within which data from an individual ground cover class are clustered.
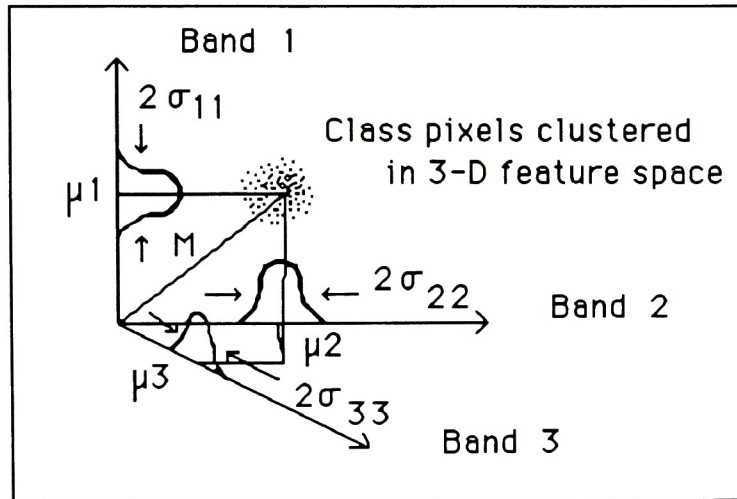
Figure   2.0-1

If all class density functions are assumed to be normally distributed in feature space, individual class statistics (i.e. mean vectors M_i and individual covariance matrices E_i) may be derived to characterize each class i.   In the feature space of Figure 2.0-1, class i mean vector M_i is defined in equations (10) and (12), and class i covariance matrix E_i is outlined in equations (11) and (13).

$$M_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \end{pmatrix} \tag{10}$$

$$E_i = \begin{pmatrix} \sigma_{11i} & \sigma_{12i} & \sigma_{13i} \\ \sigma_{21i} & \sigma_{22i} & \sigma_{23i} \\ \sigma_{31i} & \sigma_{32i} & \sigma_{33i} \end{pmatrix} \tag{11}$$

where,

$$\mu_{ki} = \sum_{l=1}^{N_i} \frac{X_k(l)}{N_i} \text{ , and} \tag{12}$$

$$\sigma_{kji} = \sum_{l=1}^{N_i} [X_k(l) - \mu_{ki}][X_j(l) - \mu_{ji}]/(N_i - 1). \tag{13}$$

$X_k(l)$ and $X_j(l)$ are the two feature values of pixel l, selected from N_i pixels for class i. Variables k and l correspond to the dimensions of each constituent of the individual class covariance matrix for class i. Mean vector components u_{1i}, u_{2i}, and u_{3i}, correspond to the projection of class i distribution onto each feature axis. Covariance terms o_{jki} represent the amount of spread of feature values from all

pixels of class i about their mean vector components between all two dimensions j, k of feature space.

2.1 Spectral, Temporal and Spatial Features

The features selected as tools to classify image data are divided into three distinct classes; spectral, temporal, and spatial. Spectral features consist of radiance measurements of image pixels in various spectral bands (i.e. measures of the electromagnetic energy reflected by and/or emitted from imaged ground targets). Temporal features are measurements of the variation in spectral response in image pixels over a given period of time. The analysis of a time-variant imaged environment is beyond the scope of this study. Finally, spatial features are measurements of spatial contrast in various spectral channels, which allow for detection of image pattern characteristics such as shape, size, texture, orientation of target objects and image context. The assumption is that the statistical dependencies between the spectral responses of neighbouring image pixels contain valuable information for classifying image data.

Since the principles of statistical classification apply for both spatial and spectral features, it is reasonable to use both types of features separately or in combination to extract pertinent image

information and increase classification accuracy, as suggested by Swain and Davis (1978).[43] Haralick et al. (1973) employed spectral features and spatial textural features to considerably improve the accuracy of satellite multispectral pattern classification relative to using spectral features alone (from 70-74% to 83.5% accuracy for classification of test data).[44] The selection of appropriate spectral and spatial features remained a function of the image data characteristics. The ECHO (Extraction and Classification of Homogeneous Objects) classifier, developed by Keltig and Landgrebe (1976), combined spatial and spectral features with homogeneity tests to increase the accuracy of supervised image pattern classification, compared to classification exercises over individual image pixels using spectral features alone.[45] Initial tests reduced the error of classification of test data from 10.5% to 5.9% for LANDSAT satellite imagery. Use of textural features alone was also compared to spectral features using the ECHO classifier, resulting in slight but consistent increases in supervised classification accuracy.

Effective use of spatial features for image pattern classification requires that there be a finite number of pixels per object to be classified. The instantaneous field of view of the receptor must be fine enough so that the number of image pixels per object is sufficient to describe the spatial characteristics of object classes. The ability to accurately classify image patterns using spectral and

spatial features is greatly increased, especially when patterns are not accurately identified using spectral features alone. This provides the analyst with greater flexibility to establish an effective feature space.

2.2   Tone and Texture for Formulation of Spatial Features

Tone and texture are fundamental concepts that allow classification of objects or regions of interest in an image, and can be characterized by means of spectral and spatial features. Tone refers to the shades of grey in a digital image while texture refers to the statistical distribution of grey tones.

Measurement of image tone consists of measuring grey-tone values over individual image pixels. Means, standard deviations and ranges of grey-tone values over entire image areas in select spectral channels also provides for similar measurements. Image tone in monochrome images is measured where one unique spectral channel is considered available and each feature constitutes a dimension in feature space.

Quantitative texture information is not directly measured by remote sensing systems, but is extracted from digital image data. Haralick (1979) reports on eight statistical approaches to measure and

quantify image texture : autocorrelations, optical and digital transforms (measures of spatial frequency), textural edgeness (the number of edge pixels per unit area), structural elements, spatial grey tone co-occurrence probabilities, grey-tone run lengths, and autoregressive models.[46]

Tonal and textural information shall be extracted from monochrome image data throughout this analysis for execution of feature selection and image data classification exercises. For clear understanding, all features used to measure tone and texture shall be referred to as "textural features". Textural features considered for use in this study are listed in Appendix A.

2.3 Textural Features Derived From Co-occurrence Matrices

Haralick discusses the texture measurement methods of Section 2.2 and suggests that the spatial grey tone co-occurrence probability approach is among the most powerful and simple statistical methods to describe texture. Weszka et al. (1976) compared textural features based on the Fourier transform and on grey-level co-occurrence probabilities, and tonal features based on local property measurements across a digitized aerial image. The most accurate

classifications were obtained using co-occurrence and local property measurement features. [47]

Grey level co-occurrences are joint probability densities of pairs of grey levels in an image texture, within a given spectral band. This approach to texture measurement is concerned with the spatial distribution of grey tones in a local area, and assumes that textural information is contained entirely in the spatial relationship between image pixels. For processing digital images, Haralick (1971) suggested two-dimensional spatial dependence of grey tones in a co-occurrence matrix for each fixed inter-pixel distance and angular relationship.[48] This matrix summarizes the frequency distribution of adjacent grey tones in a specified angular and spatial relationship over the image area of interest. The ease of derivation of this matrix is function of the subimage size, the number of quantization levels, and the spectral bands used to compute individual textural features. The ability of the co-occurrence based representation to capture texture is function of the size of the subimage region used to calculate the co-occurrence matrix, the distances separating sampled image pixels, and the matrix sampling orientation.

Figure 2.3-1 illustrates co-occurrence matrices in the horizontal (0 degree), vertical (90 degree), and oblique (45 &135 degree) image sampling orientations for unit sampling distance (d = 1) between image pixels. A 4x4 image area I is quantized to 4 grey-tone levels

where Lx is the row designate and Ly is the column designate. Indices (k,l) and (m,n) correspond to the row, column coordinates of the pixels with grey tone i and j. The grey tone co-occurrence is specified in a matrix P(i, j, d, degree) of relative frequencies with which two neighbouring image pixels, one of grey tone i and the other of grey tone j, separated by distance d and along a selected sampling orientation "degree", occur on the image area under analysis. For a given i, j, d, and degree setting, P(i, j, d, degree) is derived by carrying out analysis of grey tone levels in both directions along the established sampling orientation. The symbol # represents the number of occurrences corresponding to an established P(i, j, d, degree) setting.

Image Resolution
　　　Cells

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Gray Tone

		0	1	2	3
		0	1	2	3
Gray Tone	0	*(0,0)	*(0,1)	*(0,2)	*(0,3)
	1	*(1,0)	*(1,1)	*(1,2)	*(1,3)
	2	*(2,0)	*(2,1)	*(2,2)	*(2,3)
	3	*(3,0)	*(3,1)	*(3,2)	*(3,3)

$$P(i,j,1,0) = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \quad P(i,j,1,90) = \begin{pmatrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{pmatrix}$$

$$P(i,j,1,135) = \begin{pmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} \quad P(i,j,1,45) = \begin{pmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$P(i,j,d,0) = $ *{((k,l),(m,n)) ∈ (Ly * Lx) * (Ly * Lx)/
　　　　　　　k − m = 0, /l − n/ = d, I(k,l) = i, I(m,n) = j}

$P(i,j,d,45) = $ *{((k,l),(m,n)) ∈ (Ly * Lx) * (Ly * Lx)/
　　　　　　　(k − m = d, l − n = −d) or
　　　　　　　(k − m = −d, l − n = d), I(k,l) = i, I(m,n) = j}

$P(i,j,d,90) = $ *{((k,l),(m,n)) ∈ (Ly * Lx) * (Ly * Lx)/
　　　　　　　/k − m/ = d, l − n = 0, I(k,l) = i, I(m,n) = j}

$P(i,j,d,135) = $ *{((k,l),(m,n)) ∈ (Ly * Lx) * (Ly * Lx)/
　　　　　　　(k − m = d, l − n = d) or
　　　　　　　(k − m = −d, l − n = −d), I(k,l) = i, I(m,n) = j}

Figure   2.3-1

41

Haralick, Shanmugam and Dinstein (1973) describe a class of twenty-eight textural features which are extracted from grey-level co-occurrence matrices.[49] These features are used to classify image regions as uniform entities, where every pixel in a region is assumed to represent the same class. Several of the latter measurements, selected at random, were successfully tested for supervised classification accuracy on several types of image data using a non-parametric classifier. Photomicrograph data sets with five classes were classified with 90% accuracy. Aerial photographic data sets with eight classes were classified with 89% accuracy, and satellite imagery data sets with seven classes were classified with 82% accuracy. Eighty-three percent classification accuracy was also obtained over these data sets using a combination of textural and spectral features. Haralick and Shanmugam (1974) employed several textural features for supervised classification of image segments, extracted from LANDSAT multispectral image data, into seven classes with 70% accuracy.[50] This result compared favourably to 74% classification accuracy obtained using spectral features, consisting of means and standard deviations of grey-tone values measured over the image areas from which were derived textural co-occurrence matrices.

The spatial features proposed by Haralick et al. (1973) are all function of co-occurrence matrix sampling distance and angle. Changes in angular orientation will change the value of the feature

vectors and result in possible erroneous classification. This can be avoided by using angular averages and ranges of features for a given sampling distance as inputs to the classifier. These measures are relatively invariant under image rotation.[51]

Textural features derived from co-occurrence matrices are assigned to individual image pixels or to entire image areas. These measurements contain information on image textural characteristics such as homogeneity, grey-tone linear structures, contrast, number and nature of object boundaries, and image complexity. Assignment of textural features to individual pixels shall be carried out for all images in this study, as described in Section 4.1.1. Texture measurements based on grey-level co-occurrence matrices are inappropriate for describing the shapes of image patterns, as reported by Conners (1979).[52] A structural approach to image classification is suggested to distinguish between class shapes as referred to in Section 1.2.1. This study does not include structural features in the analysis as texture measurement is not influenced by the shape of image classes.

2.4 Other Spatial Features for Texture Analysis

Several other classes of spatial features have been reported useful for extracting textural information and for classification of image data in feature space. They include :

i) Grey-Level Difference Statistics : Weszka, Dyer and Rosenfeld (1976) propose a class of first-order statistical features derived from absolute differences between pairs of grey levels or of average grey levels.[53] The features, which are derived from the probability density functions, measure the degree of image texture coarseness and directionality.

Other useful first-order statistical features employed for texture analysis are the means and variances of local or average image property values calculated at every pixel.

ii) Grey-Level Run Length Statistics : Galloway (1975) proposes a class of spatial features computed over regions which reflect runs of consecutive pixels of identical grey-tone value.[54] These features are referred to as grey-level run-length statistics and are indicative of individual image run grey-tone value, length, and direction. Galloway demonstrated the effectiveness of these features, combined with textural feature data sets proposed by Haralick, for increased classification accuracy. These measures

remain effective if image noise does not influence the values of measured image grey tones.

Classification of image textures by Weszka, Dyer and Rosenfeld (1976) using sets of co-occurrence matrix, run length, and grey-level difference statistics features resulted in varying but intuitively acceptable degrees of accuracy in the 70th to 80th percentile for classification of LANDSAT multispectral image data.[55] These features were equalized in orientation and sampling size for comparison. Run-length statistics were eventually abandoned due to their sensitivity to image noise. Single sets of the remaining features did equally well where the best feature in each set obtained 75% classification accuracy. The sampling directionality significantly affected classification accuracy; classification accuracy increased dramatically when the best feature pair in each set correctly classified 93% of the image data. Pattern coarseness or sampling size was reported to be generally more important than directionality for discriminating between particular textures.

iii) Textural Edgeness : Sutton and Hall (1972) propose the gradient of textural edgeness (i.e. the amount of edge per unit image area) dependant on the distance between sampled pixels.[56] The proposed feature was applied by the authors with classification accuracy in the 80th percentile. Rosenfeld (1975)[57] proposes a similar

approach to measuring textural edgeness by computing the average value of the quick Robert's gradient over the image of interest, as given by Rosenfeld and Thurston (1971).[58] Duda and Hart (1973) also propose the Sobel edge operator to obtain the smoothed gradient output of an edge in a given direction.[59]

iv) Composite Features : Weszka, Dyer and Rosenfeld (1976) suggest features derived from other sets of features for texture analysis.[60] Means and standard deviations of various textural features over all directions for all sampling distances are typical examples. Composite features performed as well as the sets of features from which they were derived.

v) Texture Transformations : Hsu (1978)[61] and Irons and Peterson (1981)[62] propose generalizing the grey-tone co-occurrence textural feature extractor to a textural transform mode. Grey-tone co-occurrence features derived over an entire image produce textural features at a coarser resolution than the original image. By doubling or tripling the computation time required to determine grey-tone co-occurrence matrices, it is possible to provide a textural transform that preserves resolution. This information is derived from both local and global grey-tone co-occurrences

Other spatial features exist for texture analysis such as measurement of colour disposition in multispectral images.

Features which were selected for treatment of individual images throughout this study are included in Appendix A-1.

2.5   Feature Space Dimensionality

Image classification accuracy is dependant upon the nature and the number of features.   It would seem that classification accuracy would increase significantly if one was to add additional features to the feature space.   Estes et al. (1983) state that this is only true if the additional features do not contain redundant information.[63] This is rarely true for remotely sensed data since certain features (especially spectral features) are highly correlated.   To achieve precision in classification, it is important to define a set of meaningful features which accurately describe the information contained in an image.

The analyst must realize that the costs of classification with large numbers of features is often excessive.   For the Maximum Likelihood classifier proposed in this study, the cost associated with feature numbers climbs quadratically, as stated by Richards (1986).[64] The analyst must be able to identify and use the most important features.

Feature selection reduces the dimensionality of the feature space by removing those with little or no discriminatory ability. Combining or transforming features to ensure increased separability between classes is referred to as feature reduction.

2.5.1   Feature Selection

Many feature selection algorithms will increase the accuracy and efficiency of the classification process. Regardless of their operating methodologies, they share one common goal: to reduce the number of features required to accurately classify patterns. Feature selection techniques such as Mahalanobis feature selection, intraset and interset Euclidean feature selection, Karhunen-Loeve feature selection, Fisher feature selection, minimum entropy selection in normal distribution and equal covariance, and divergence feature selection in normal distribution are but a few of the more popular methods available, as reported by Xuan (1984).[65]

To test the discriminatory ability of feature subsets to accurately classify image data, several feature selection methods attempt to minimize the Bayesian probability of misclassification. Others employ probabilistic distance or statistical separability measures between classes in feature space, as reported by Morgera and Datta (1984).[66] The approach chosen for this study measures the

mathematical separability between all image classes in feature space for a reduced set of features, in terms of Mahalanobis-based distance measurements. This method was perfered due to the availability of software and to inherent difficulty of measuring the probability of misclassification (or Bayes) error. If class separability is not reduced by removing features from the original set, these features can be conveniently discarded from the analysis as they offer little or no discriminability.

Chen (1973) compared distance related feature selection approaches for gaussian classes with equal means and covariances. He recommends statistics such as the Bhattacharrya (or Jeffries-Matusita) distance and divergence measures as discriminators.[67] The ability of selected features to correctly classify patterns may be quantitatively assured when using statistical image pattern classifiers. The Bhattacharrya distance is a measure of the separation of class probability distributions in feature space based on the average separation distance. Transformed divergence indicates class separability based upon the degree of overlap. Chen elaborates upon the Bayes' risk for the Bhattacharrya measure and establishes useful upper and lower bounds. Richards (1986) emphasizes the effectiveness of both the Bhattacharrya distance and the transformed divergence for use with the Maximum Likelihood classifier when the multivariate class probability distributions are assumed normal.[68] The transformed divergence statistic is

reportedly more suitable and cost-effective to implement than the Bhattacharrya distance if a large number of feature subsets are to be assessed from a set of features. These feature selection techniques were not implemented in this study due to the unavailability of source code.

Schott et al. (1988) proposed an optimal feature selection technique based on class separation metrics which were generated from individual weighted-class divergence matrices.[69] This method of feature selection is preferred in this study due to computational efficiency, accuracy, and appropriateness for multivariate normally distributed class probability distributions in feature space. The Gaussian Maximum Likelihood Classification theory (Section 1.6) is the basis for this technique, and enables derivation of Mahalanobis-like distance measures between individual classes in feature space.

For N image classes characterized by gaussian distributions in a K-dimensional feature space, the NxN class divergence matrix DM is constructed as shown in Figure 2.5.1-1. This matrix allows the analyst to observe class separability.

$$DM = \begin{bmatrix} d11 \ d12 \ d13 \ \ldots\ldots\ldots \ d1n \\ d21 \ d22 \ d23 \ \ldots\ldots\ldots \ d2n \\ d31 \ d32 \ d33 \ \ldots\ldots\ldots \ d3n \\ \vdots \qquad\qquad\qquad \vdots \\ dn1 \ dn2 \ dn3 \ \ldots\ldots\ldots \ dnn \end{bmatrix}$$

Figure   2.5.1-1

Each value of d_{ij} corresponds to a Mahalanobis-like distance measure between classes i and j and is defined in equation (13) :

$$d_{ij} = (Mi - Mj)^t * E_i{}^{-1} * (Mi - Mj), \qquad (13)$$

where :

$(M_i - M_j)^t$: transpose of difference between mean vectors

of classes i and j in K dimensional feature

space, and

$E_i{}^{-1}$: inverse of KxK individual covariance matrix E_i

for class i.

Equation (13) is different from the true Mahalanobis distance of equation (9) as it includes individual class covariance matrices.   The

term Ai of equation (7) is also excluded, despite not being a constant over all classes when using individual class covariance matrices, and should be considered in future studies.

To render computation more efficient, pooled class covariance matrices E_p may be substituted for individual covariance matrices E_i in (13) when the determinant of all E_i are equal. Use of pooled class covariance matrices decreases the accuracy but increases the efficiency of the feature selection process.

Optimal features are selected on the basis of a class separation metric Z which is generated from a weighted version of the divergence matrix WDM, as shown in Figure 2.5.2-1. The class separation metric used to select optimal features is the square root of the weighted sum of all Mahalanobis distances over all classes i and j in feature space and is defined in equation (14) :

$$WDM = \begin{Bmatrix} w11d11 \ w12d12 \ ....w1nd1n \\ w21d21 \ w21d22...w2nd2n \\ w31d31 \ w32d32....w3nd3n \\ \vdots \qquad\qquad \vdots \\ wn1dn1 \ wn2dn2 \ ....wnndnn \end{Bmatrix}$$

Figure   2.5.2-1

and

$$Z = \left/ \sum_{j=1}^{K} \sum_{i=1}^{K} w_{ij} * d_{ij} \right/^{**} 0.5, \qquad (14)$$

where :

d_{ij} : Mahalanobis-like distance from equation (13), and

w_{ij} : weight factor assigned to d_{ij}.

The values of the weights that characterize each component of the weighted divergence matrix are dictated by scenario specific requirements and are set equal to 1.0 for the purpose of this study. Optimal feature selection will maximize the class separation metric

53

Z and the overall separation between classes in feature space. This ensures maximum classification accuracy and minimum classification error.

This feature selection process considers each combination of M desired features from an initial data base of K features, and selects the optimal combination (i.e. the combination that maximizes Z). From an initial feature space of dimension K, all possible combinations of features are assembled for best " M- " dimensional feature spaces (where M = 2, 3, ..., K) and corresponding maximum divergence statistics are established. For each best " M- " dimensional feature space, the combination of features with maximum value of Z is selected. By establishing a threshold value for increases in the value of Z for increments in the dimensionality of optimal feature space, or by defaulting to maximum value of Z, a final optimal feature space is derived.

Another approach to feature selection employed in this study reduces the dimensionality of the initial feature space by manipulating correlation measures derived from pooled class covariance matrices. Features with a high degree of correlation contain redundant information and can be removed from further analysis. For a K-dimensional feature space and N classes identified therein, a KxK pooled class covariance matrix is derived from N individual covariance matrices to establish a correlation coefficient

matrix.   Pooled class covariance matrices are derived from equations (11) and (12), and correlation coefficients CC between all features i and j (where i and j = 1, 2, ..., K) of the initial feature space are derived as per equation (15) :

$$CC_{ij} = \sigma_{ij} / (\sigma_{ii} * \sigma_{jj})^{**0.5} \qquad (15)$$

Once formulated between all cases i and j, a correlation coefficient matrix is established of dimension KxK.   Considering correlation coefficients by column within the correlation coefficient matrix, the correlation coefficients which are larger than an established threshold are grouped into independent subsets.   Their rows are then deleted from consideration from the correlation coefficient matrix. This analysis is carried out until all matrix columns are exhausted. From each subset, the largest correlation coefficient is extracted and its first dimension is a feature of the *initial optimal feature space.*

Figure 2.5.3-1 illustrates an example of this method of feature selection for a correlation coefficient matrix of dimension 4x4.   The initial feature space consists of 4 features and a threshold value of 0.6 is established.   By manipulating the correlation coefficient matrix as explained above, two uncorrelated features are extracted from the initial set of four.

Step #1

$$\begin{pmatrix} 1.0 & 0.4 & 0.3 & 0.7 \\ 0.4 & 1.0 & 0.8 & 0.2 \\ 0.3 & 0.8 & 1.0 & 0.9 \\ 0.7 & 0.2 & 0.9 & 1.0 \end{pmatrix}$$

1) Correlation Coefficient

Matrix of dimension 4x4

Threshold = 0.6

Step #2

$$\begin{pmatrix} \boxed{1.0} & 0.4 & 0.3 & 0.7 \\ 0.4 & 1.0 & 0.8 & 0.2 \\ 0.3 & 0.8 & 1.0 & 0.9 \\ \boxed{0.7} & 0.2 & 0.9 & 1.0 \end{pmatrix}$$

2) Consider Column #1

$CC_{11} = 1.0$ ✳

$CC_{41} = 0.7$

Highest CC = CC_{11}
First dimension : Feature #1

Step #3

$$\begin{pmatrix} & & & \\ 0.4 & \boxed{1.0} & 0.8 & 0.2 \\ 0.3 & \boxed{0.8} & 1.0 & 0.9 \\ & & & \end{pmatrix}$$

3) Consider Column #2

$CC22 = 1.0$ ✳

$CC32 = 0.8$

Highest CC = CC_{22}
First dimension : Feature #2

Step #4

4) Initial Optimal Feature Space : Features #1,2

Figure   2.5.3-1

All initial optimal features are subjected to the feature selection analysis proposed by Schott et al. (1988) to derive the final optimal feature space.

Instead of computing a pooled class covariance matrix for feature selection as per equation (8), it is suggested that such matrix be derived directly from all observations (i.e. consider all pixels are from a same class). This provides for a more accurate estimator of pooled class covariance when the determinant of individual class covariance matrices are unequal. Although this method was not adopted, it is suggested for consideration in future studies.

The combined method of deriving correlation coefficients from pooled class covariance matrices and feature selection techniques proposed by Schott et al. are used to derive final optimal features for individual treatments on separate images in this study.

2.5.2   Feature Reduction

Feature reduction transforms image data to a new set of coordinates in feature space to increase the separability of classes. The number of features remains the same, and the data information is concentrated in the first transformed coordinates. Transformed coordinates with little or no discriminatory ability are removed

from the analysis. This transformation increases the speed of the classification process and results in data of a superior quality. Although many feature reduction algorithms exist, the most commonly employed in remote sensing are principal components analysis and canonical analysis.

The principal components transformation creates a new feature space of equal dimension and less correlation. The transformed coordinates which contain small amounts of variances over image data are discarded as they contribute little to the separability of image classes. Crist and Kauth (1986) demonstrated that six reflective spectral bands of LANDSAT TM were effectively transformed to three bands of information that contained 95% of the tonal variance in the original data. [70]

If separability after principal components is inadequate, the canonical transformation is a possible alternative for feature reduction. This method rotates feature axes to ensure maximum class separability. Axes are determined to maximize class separation and minimize class variability. This method of feature reduction is much more sensitive to class structure than the principal components transformation. Richards (1986) offers a detailed mathematical description of these feature reduction techniques.[71]

Feature selection techniques are chosen to be implemented in this study as opposed to feature reduction algorithms. It is judged that although feature reduction may increase classification accuracy they also increase associated computational costs. This is a direct result of algorithms (such as principal components) that maximize the amount of data over a minimum number of variables by creating new features from linear combinations of existing features. Principal component analysis also requires that all features be derived to allow for feature reduction. Finally, feature reduction algorithms usually provide uninterpretable combinations of original features. Feature selection algorithms provide for increases in both classification accuracy and effectiveness.

3.0   Method of Analysis

In this study, final optimal feature spaces are established for use on several digitized monochrome images.   The ability to discriminate is evaluated by classifying sequences of image test and training data using a Gaussian Maximum Likelihood Classifier.   The effect on classification accuracy from textural feature parameter settings is assessed for each.   The procedure employed is as follows:

1) Each monochrome image to be analysed is extracted from an individual spectral channel of false-colour infra-red film.   The image is digitized to 256 grey levels and 512x512 pixels.   The selected channel provides for maximum visible texture across all classes.

(2)   Each image is preprocessed to correct for distortions and/or degradations.   Preprocessing operations include geometric manipulation, radiometric correction, and contrast manipulation (grey-level thresholding, level slicing, and contrast stretching).   The histograms of all images are flattened to ensure that all grey levels are equally present and remove the effects of unequal overall brightness and contrast.   Otherwise, such effects might dominate the classification and disrupt image classification accuracy. Histogram level slicing is performed on all images to remove

unwanted image information (i.e. such as noise) and to isolate image patterns.

(3) For each image, the size of the image sampling window and the interpixel sampling distance are established to derive feature measurements for individual training pixels. Caution is exercised to ensure that all textural features measure the textural information of all image classes. Class mean vectors are derived and individual and pooled class covariance matrices are generated. From these statistics, correlation coefficient matrices, weighted class divergence matrices and class separation metrics are calculated. Uncorrelated features are extracted to establish initial and final optimal feature spaces in accordance with the feature selection techniques proposed for use in this study. Final optimal features allow for subsequent best performance of a trained Bayes Optimal classifier.

(4) The training data is used to train a Gaussian Maximum Likelihood classifier for each image. Supervised classification of image training and test data is performed to assess classifier accuracy and efficiency, to establish the validity of employing textural features for classifying monochrome imagery, and to determine the ability of the derived final optimal feature space to separate between image classes. Gaussian Maximum Likelihood classification is also performed on textural image data from several image

treatments using ERDAS software. This ensures selection of unbiased test pixels to evaluate the appropriate classifier.

(5) The ability of final optimal feature spaces for classification is assessed for several images by repeating the analysis on identical monochrome imagery with varying resolution and orientation. Textural feature parameters are repeated to establish their effects on classification accuracy.

A flowchart of the method for all images is shown in Appendix E. Each block in this flowchart reflects the use of specific source code for deriving the required results. These programs are listed in the following section.

3.1 Implementation of Software

All code is written in Fortran 77 and is available from the Imaging Science department. Five main programs are used to perform analysis as listed in the paragraphs which follow.

Program #1 derives textural features for individual pixels in selected image windows. Parameters such as interpixel sampling distance, number of image quantization levels, and image dimension

areas are established to capture the desired textural feature measurements.

Program #2 reduces the dimensionality of the initial image feature space by analyzing pooled class covariance matrices, and by deriving correlation coefficients between all dimensions included therein. For an established absolute threshold (between 0.0 and 1.0, and chosen as 0.6 for this study), the correlation coefficients that share a common second dimension are grouped in individual sets, from which are derived initial optimal textural features. All such features possess a maximum correlation coefficient within their individual set and make up the image's *initial optimal feature space.*

Program #3 further reduces the dimensionality of the initial optimal feature space by establishing the "n" best feature combinations from its components (i.e. best 2 features for pixel classification, best 3 features, ... , best n features). For specific dimensionality, the best combinations of features provide for maximum global separability among all image classes within the boundaries of the individual feature spaces. This quantity is measured in terms of an expressed statistic entitled class separation metric Z. From the best n feature combinations, a final optimal feature space is extracted which provides for an increase of no less then 1% in Z between consecutive best feature combinations or for a maximum value in the latter measure. The strict value of the threshold 1%

was established by the analyst to maximize feature selection accuracy. Figures 3.0.1-1 and 3.0.2-1 are typical examples of final selected features, based on the class separation metric. In Figure 3.0.1-1, 8 final optimal features are selected from an initial optimal feature space of 12, based on measures of 1% increase in the Z statistic. In Figure 3.0.2-1, 7 final optimal features are selected from an initial optimal feature space of 17 features, based on maximum value of Z statistic.
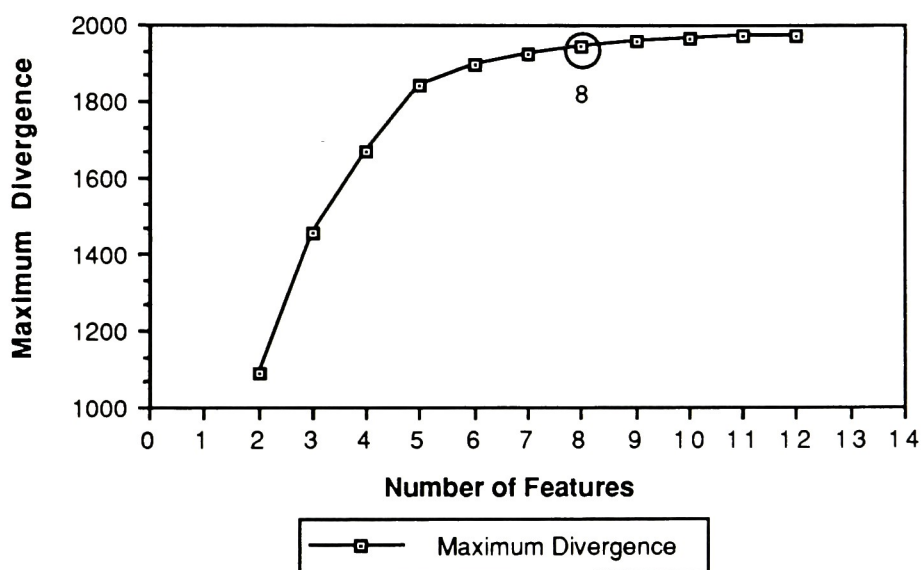
Figure 3.0.1-1
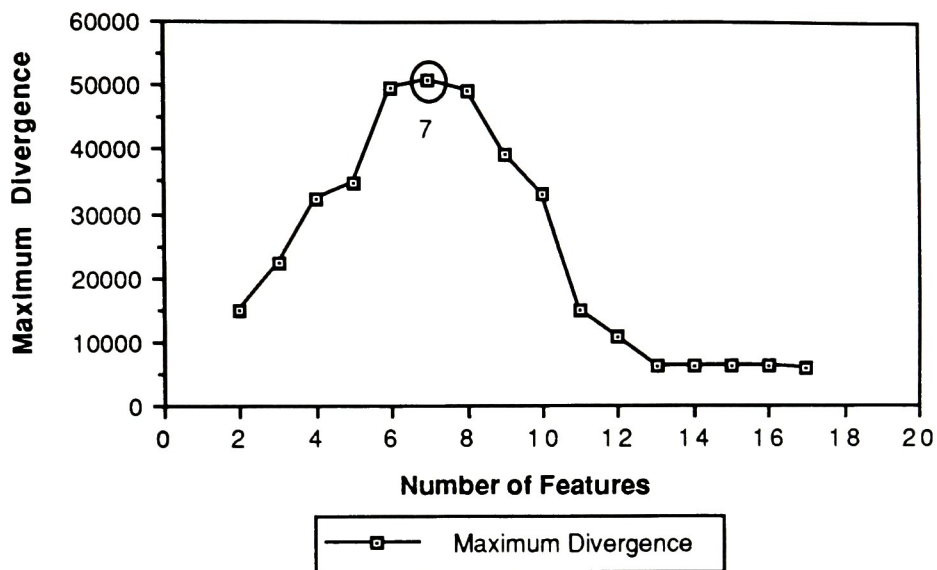
Figure   3.0.2-1

The even shape of the curve in Figure 3.0.1-1 is result of feature selection by discriminant analysis when all class covariance matrices are non-singular in nature. The uneven shape of the curve in Figure 3.0.2-1 is result of derivation of the Moore-Penrose generalized inverse of singular covariance matrices.

Program #3 also derives individual class statistics (i.e. class mean vectors, pooled class covariance matrices, and individual class covariance matrices) in the feature spaces.

Program #4 classifies image training and test data into established classes. For each treatment, the class statistics derived from

65

Program #3 assist in establishing accurate discriminant functions in final optimal feature space which serve to classify individual image pixels. Global classification accuracy is used to measure the separability of classes into individual and distinct distributions, and to assess the ability of the classifier to accurately discriminate between dissimilar image pixels.

Program #5 consists of several subprograms which perform display of textural feature images, and reformatting of textural feature data files to reflect the dimensions of optimal feature spaces.

Programs #1 through #5 are used for all images. As many of these programs require significant computer processing time, all programs were usually submitted in batch mode.

4.0   Results from Analysis

This section will discuss preprocessing operations and will present results of the derivation of final optimal feature spaces, and of the classification of training and test data for all image analyses.

4.1   Processing of Digital Images

Three aerial scenes were selected for analysis, entitled **Urban**, **Field**, and **Forest** Each scene had been recorded on false colour infra-red film, and was characterized by multiple ground cover class content and individual class textures. All images were digitized to 512x512 pixels over 256 grey levels in the red spectral channel. Prior to testing and evaluation, the histogram of each digital image was equalized and level slicing preprocessing was performed to ensure good image quality.

The assumption of normal distribution of class probability density functions in final optimal feature space was verified for all image analyses. Histograms of final optimal feature values were plotted for all training pixels over each class and all distributions were characterized by a skewed and familiar unimodal normal shape. Section 1.4 provides further discussion.

The processed images were analysed to establish optimal feature spaces for maximum classification accuracy of digital image data,

to validate the process by which they were constructed, and to determine the effects on classification accuracy due to varying image sampling window size, number of image quantization levels, image magnification, and image rotation. The image data were classified in supervised fashion using a Gaussian Maximum Likelihood Classifier in final optimal feature space.

4.1.1   Image Class Windows and Image Sampling Windows

All analyses involved selecting image class and image sampling windows across class areas to derive textural features. The size and location of all windows were varied to capture appropriate pixels to train classifiers. Image sampling windows were scanned across and within the boundaries of image class windows in a row by row fashion and were of an odd number of pixels in square dimension to facilitate deposition of results. Textural features were attributed to pixels at the center position of sampling windows.

For all analyses, the sizes of the class and sampling windows were set to encompass uniformly textured areas, to provide discriminant measurements, to capture a sufficient number of training samples for accurate estimation of class distributions, and to satisfy computing requirements. Despite attempts to process images in accordance with the above criteria, image restrictions (i.e. class sizes and shapes) and code limitations (i.e. dimensions of arrays)

resulted in frequent exceptions to these criteria. These exceptions resulted in capture of insufficient number of training pixels for small classes. Selection of training pixels from multiple areas of common texture was attempted to minimize their effects. Figure 4.1.1-1 illustrates placement of image class and sampling windows across image textures.

IMAGE CLASS AND IMAGE SAMPLING WINDOWS

IMAGE CLASS WINDOW

IMAGE SAMPLING WINDOW

DIGITIZED MONOCHROME IMAGE
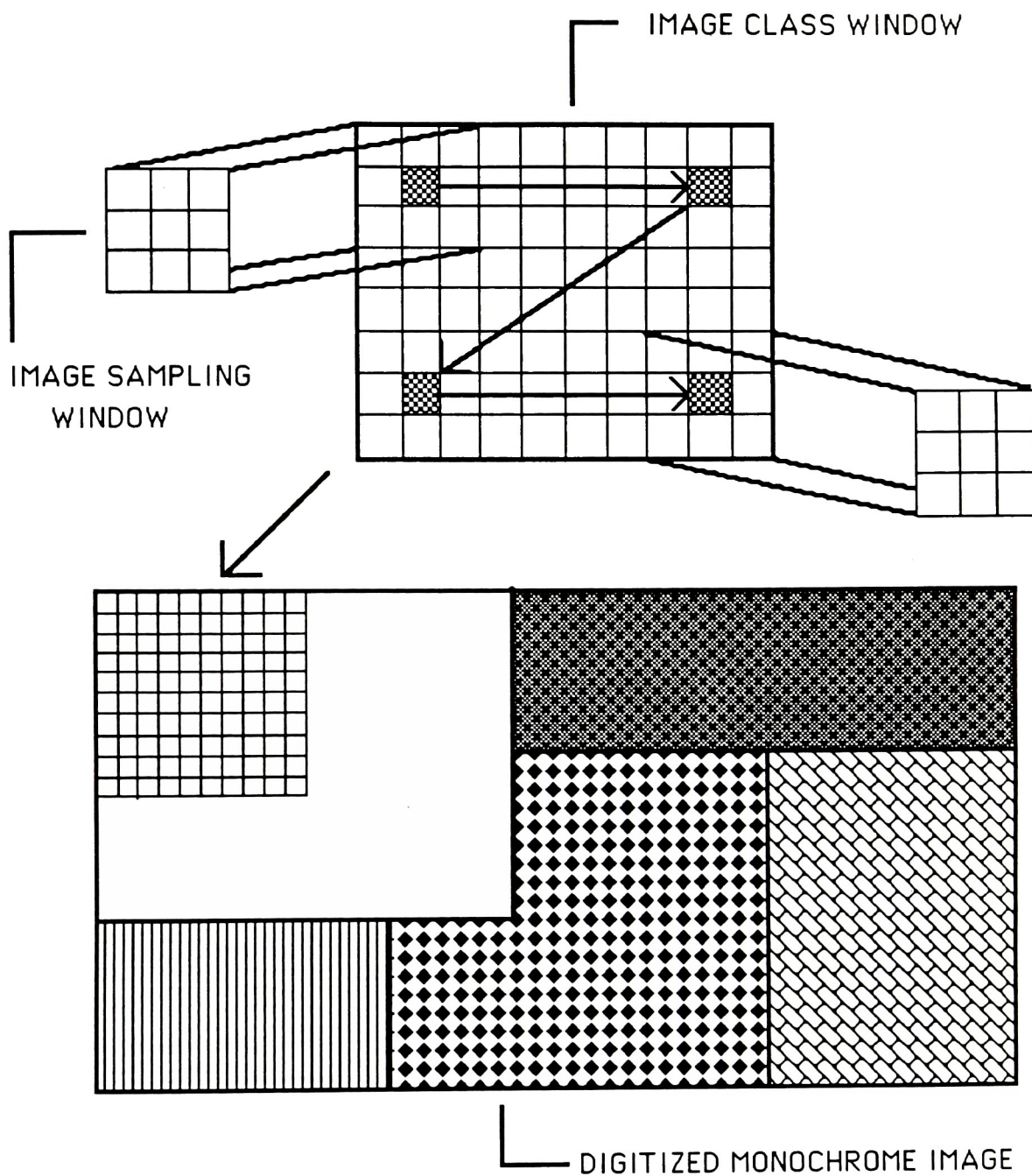
Figure   4.1.1-1

4.2   Processing of Image Urban

Eleven ground cover classes were identified within image Urban entitled : 1) light grey water, 2) medium grey water, 3) dark grey water, 4) residential, 5) light grey forest, 6) dark grey forest, 7) orchard, 8) light grey grass, 9) medium grey grass, 10) dark grey grass, and 11) highway.   The majority of these classes displayed overall uniform texture with exception to class 4, where texture varied from region to region.   Image Urban is shown in Figure 4.2-1.
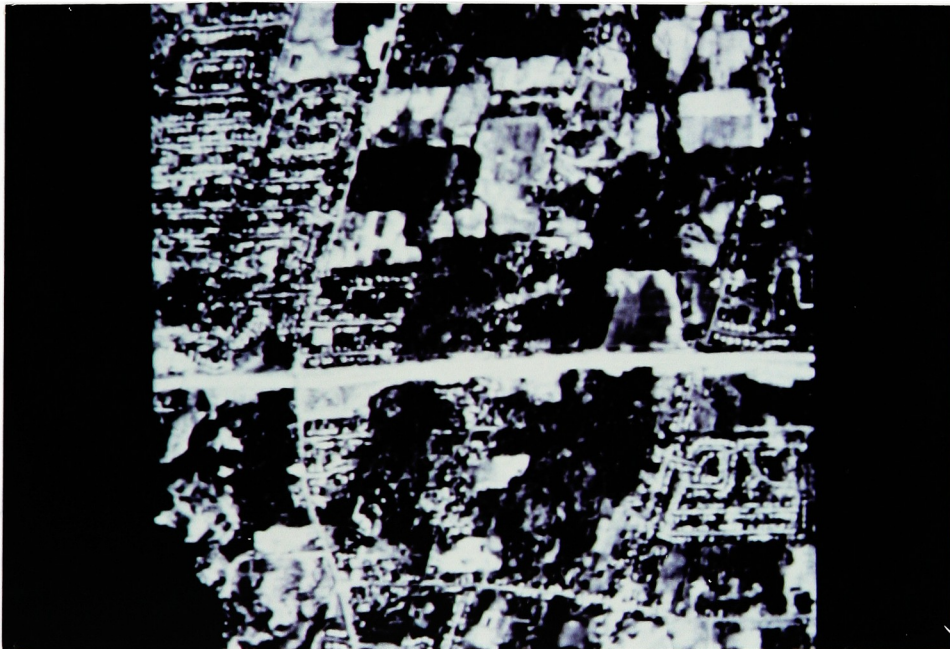
Figure   4.2-1

4.2.1   Analyses 162, 164, 322, 324

Image Urban was processed in 4 ways entitled : 162, 164, 322, and 324 as listed in Table 4.2-1.   Analyses (162, 164) and (322, 324) displayed the effects of varying image sampling window size and constant number of image grey tone quantization levels on classification accuracy.   Analyses (162, 322) and (164, 324) displayed similar effects from varying number of image grey tone quantization levels and constant image sampling window size.

TABLE   4.2-1

Analysis Parameter Settings for Analyses 162, 164, 322, 324 over Image Urban

Analysis : Parameters	162	164	322	324
Quantization Levels	16	16	32	32
Sampling Window Size	3*3	5*5	3*3	5*5
Interpixel Sampling	1	1	1	1
Degrees of Rotation	0	0	0	0

Nineteen image class windows were selected across the eleven established image ground cover classes within the image. Forty-five textural features were measured over 2741 training pixels for analyses 162 and 322, and over 1875 training pixels for analyses 164 and 324. All textural features used in analyses over image Urban are listed in Appendix A-1. The distribution of training pixels per image class for each analysis is listed in Table 4.2-2. Limited numbers of training pixels per class were due to the small size of classes compared to the size of image class and sampling windows.

TABLE 4.2-2

Number of Training Pixels per Class for Analyses 162, 164, 322, 324 on Image Urban

Image Classes	Analysis 162/322	164/324
1) light grey water	63	35
2) medium grey water	72	40
3) dark grey water	35	15
4) residential	805	613
5) light grey forest	330	216
6) dark grey forest	232	154
7) orchard	323	229
8) light grey grass	117	77
9) medium grey grass	136	90
10) dark grey grass	408	284
11) highway	220	122

For each analysis, class individual matrices and a class pooled covariance matrix of dimension 45x45 were derived from the initial training data. From each pooled covariance matrix derived as per equation (8), a feature correlation coefficient matrix of equal dimension was extracted. In accordance with the correlation coefficient matrix processing technique for feature selection explained in Section 2.5.1, a threshold of 60% was established, above and beyond which all correlation coefficients sharing a common second dimension were grouped. An individual set of uncorrelated initial optimal features was obtained for each analysis; consisting of 11 features for analysis 162, 17 features for analysis 164, 12 features for analysis 322, and finally 17 features for analysis 324. For analysis 162, the initial optimal features were measured over all image pixels and textural feature images were derived. Initial optimal feature spaces for all analyses are listed in Appendix A-2 while textural feature images for analysis 162 are displayed in Appendix A-10.

The approach suggested in Section 2.5.1 for deriving a more accurate covariance matrix over all classes for establishment of initial optimal features was carried out for all analyses. By considering all training pixels belonging to a same class for each analysis the latter matrix was derived. It was discovered that this matrix yielded dissimilar initial optimal features from those derived from the pooled covariance matrix used in this study when using the same feature selection technique (see Figure 2.5.3-1). This investigation was also carried out for all analyses over images Field and Forest

74

and similar conclusions were made. Further investigation is recommended.

For each analysis, the data file containing the original training data was reformatted to contain the features of the initial optimal feature space. From each data file class statistics in initial optimal feature space, subsets of optimal feature combinations (i.e. best 2 features for classification, best 3 features for classification, ... , best n features for classification), and associated measures of maximum divergence Z were derived. By establishing a limit of no less than 1% increase in maximum divergence statistic between best feature combinations, or defaulting to its maximum value, final optimal feature spaces were extracted. Figures 4.2-2 through 4.2-5 show the dimensions of final optimal feature spaces in terms of Maximum Divergence vs Number of Features in Final Optimal Feature Space. Circled data points represent these dimensions. Appendix A-2 lists final optimal features. Numbers of features in initial and final optimal feature spaces are listed in Figure 4.2-6. Best feature combinations for each analysis are listed in Appendix A-9.
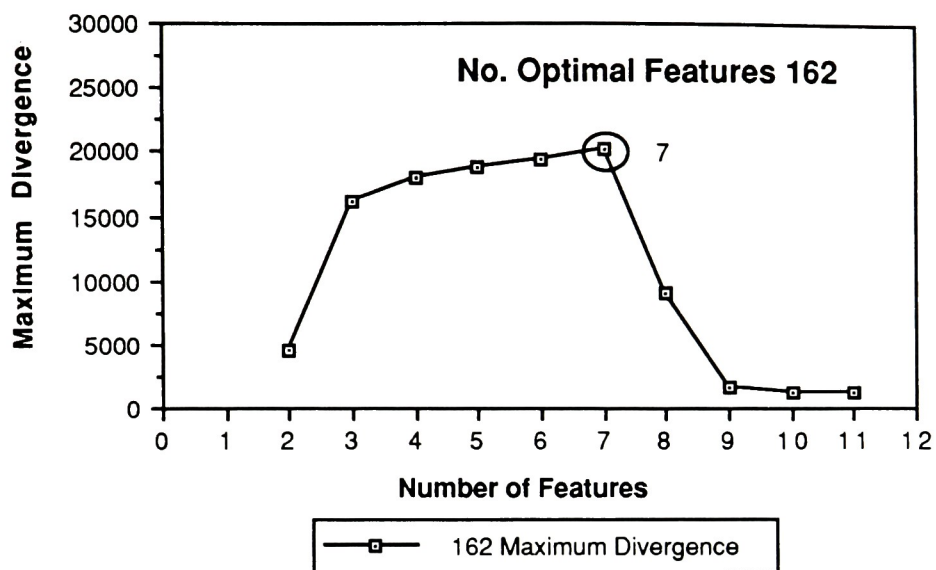
Figure   4.2-2
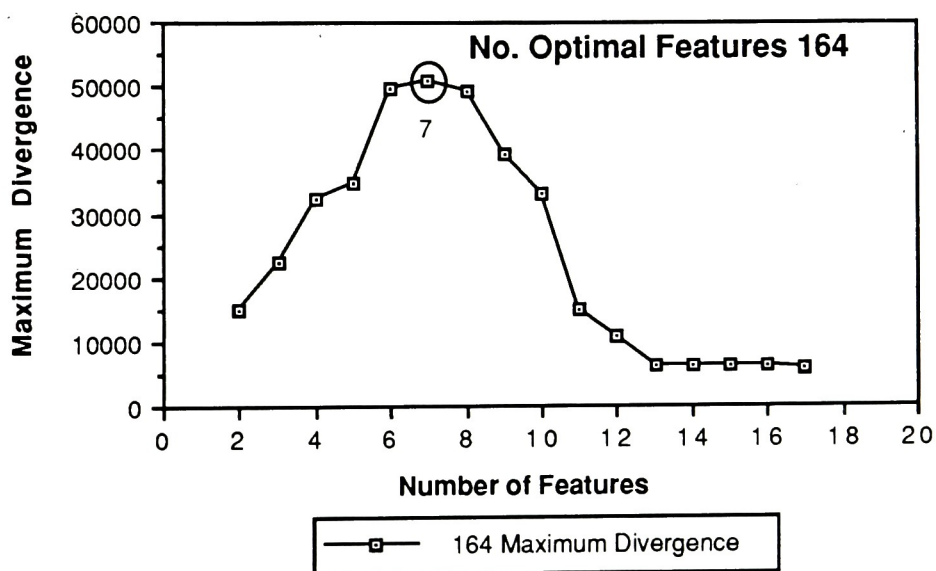
Figure   4.2-3
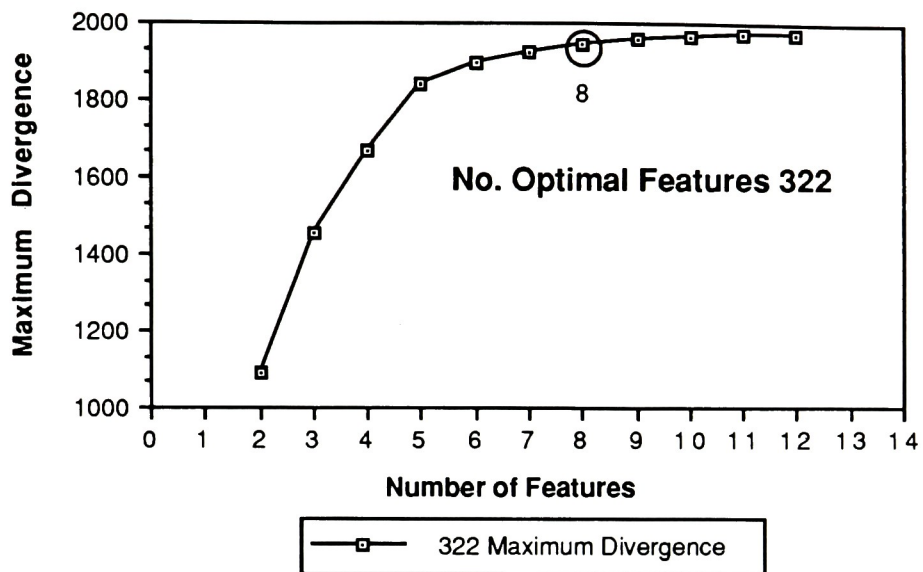
Figure   4.2-4
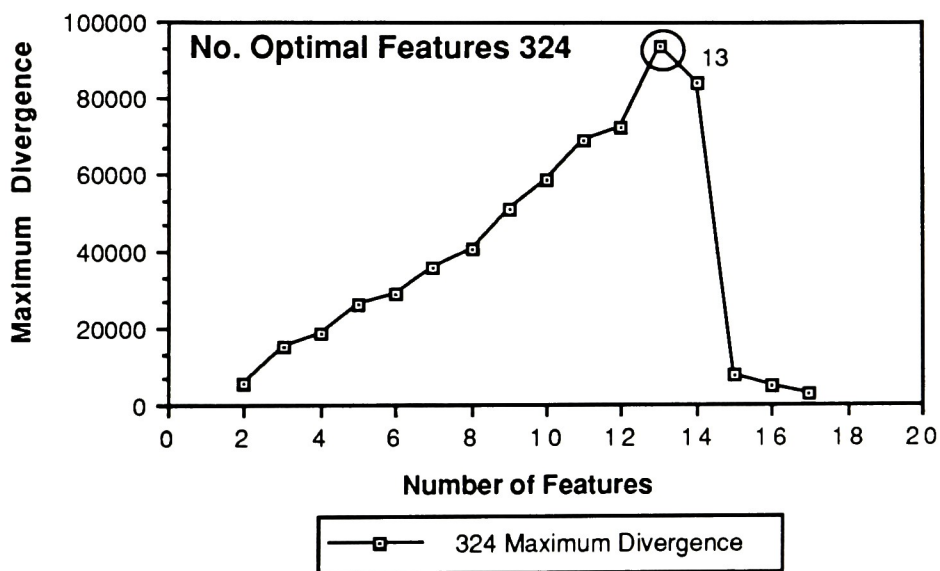
Figure   4.2-5

INITIAL AND FINAL OPTIMAL FEATURE SPACES

```
┌──────────┐  ┌──────────┐  ┌──────────┐  ┌──────────┐
│   162    │  │   164    │  │   322    │  │   324    │
│          │  │          │  │          │  │          │
│ 45 feat. │  │ 45 feat. │  │ 45 feat. │  │ 45 feat. │
└──────────┘  └──────────┘  └──────────┘  └──────────┘
     ↓             ↓             ↓             ↓
```

INITIAL OPTIMAL FEATURE SPACE

11 feat.   17 feat.   12 feat.   17 feat.

```
     ↓             ↓             ↓             ↓
```

FINAL OPTIMAL FEATURE SPACE

7 feat.   7 feat.   8 feat.   13 feat.
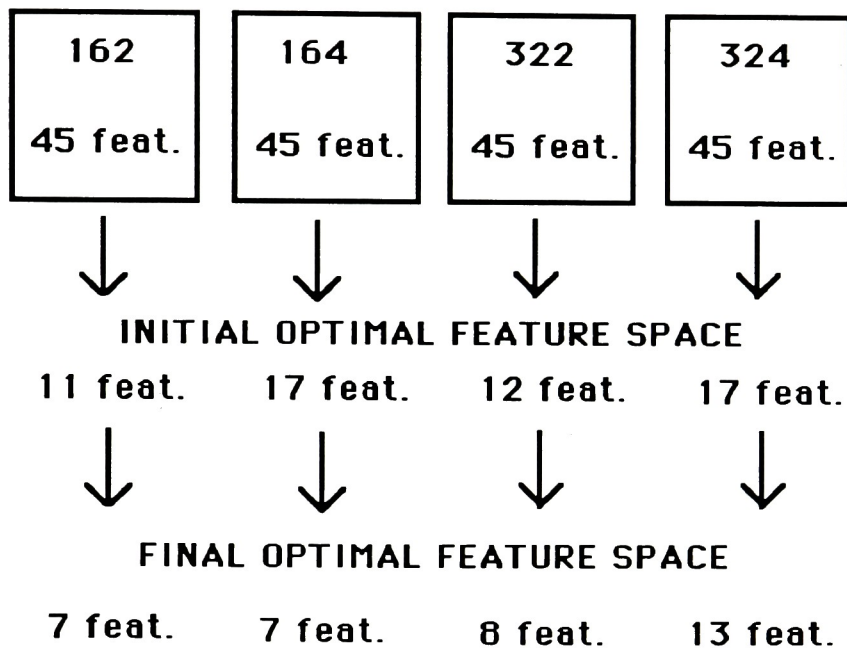
Figure   4.2-6

The initial training data were processed to include features of final optimal feature space and class statistics were derived.   Image training and test data were classified using individual class covariance matrices.   The image test data consisted of thirty pixels per class, selected at random by the analyst, across the face of each of the 11 image classes.   Results of classification in terms of percentage of training and test data accurately classified are listed in Appendix A-3 and in the plots of Figures 4.2-7 through 4.2-10.

The abbreviation GCA in the following figures corresponds to "Global Classification Accuracy". The term "Roch" in all plots relates to analyses on image Urban.
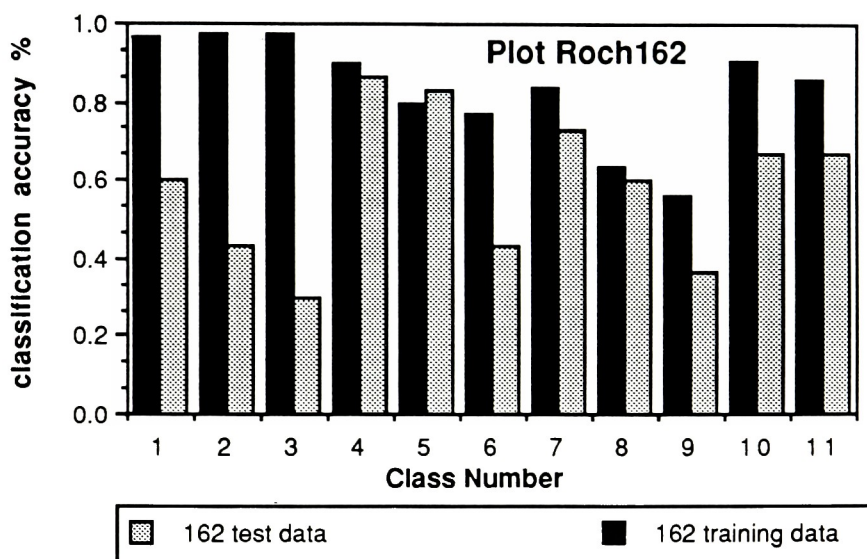
GCA Training Data : 83.4%   GCA Test Data : 59.0%

Figure   4.2-7

Plot Roch164
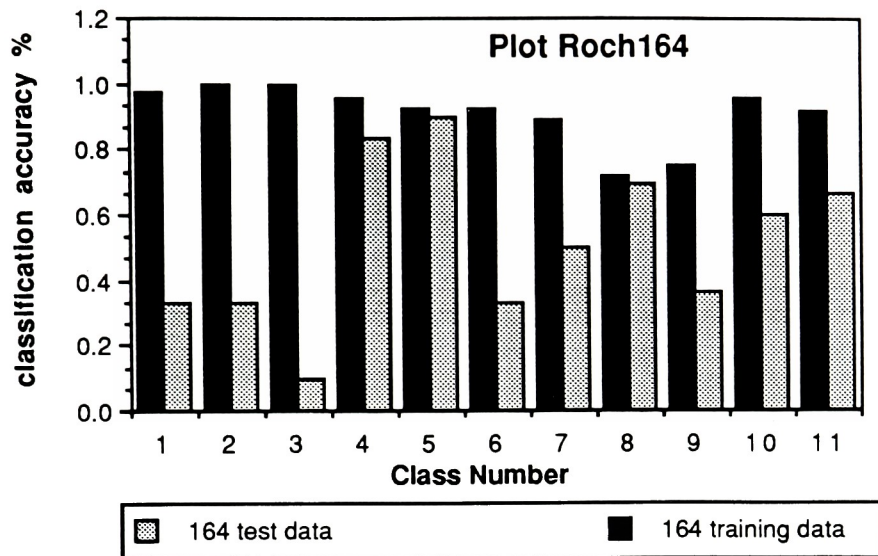
GCA Training Data : 91.1%   GCA Test Data : 51.5%

Figure   4.2-8

Plot Roch322
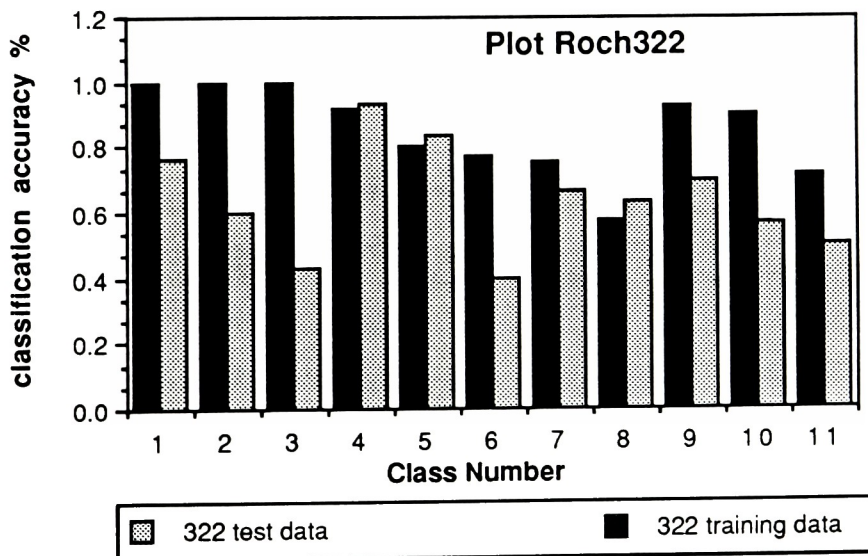
GCA Training Data : 85.1%   GCA Test Data : 63.9%

Figure   4.2-9
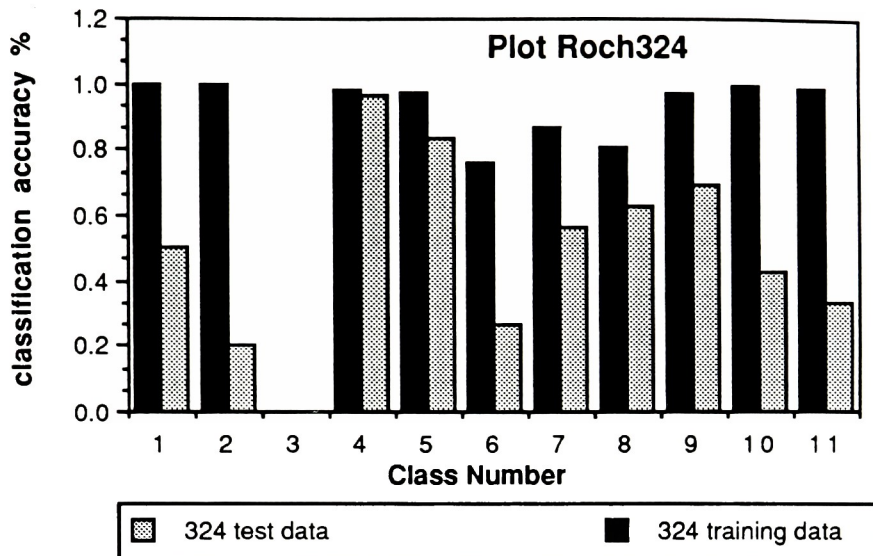
Figure 4.2-10

GCA Training Data : 85.0%   GCA Test Data : 49.4%

4.2.2 Analyses 261, 461, 223, 423

To illustrate the "next-to optimal" characteristic of final optimal feature spaces for analyses 162, 164, 322, and 324, analyses 261, 461, 223, and 423 were carried out. Final "sub-optimal" feature spaces were derived from feature sets which excluded final optimal features derived from initial analyses 162, 164, 322, and 324, and image training and test data were classified. Figure 4.2.-11 illustrates this process.
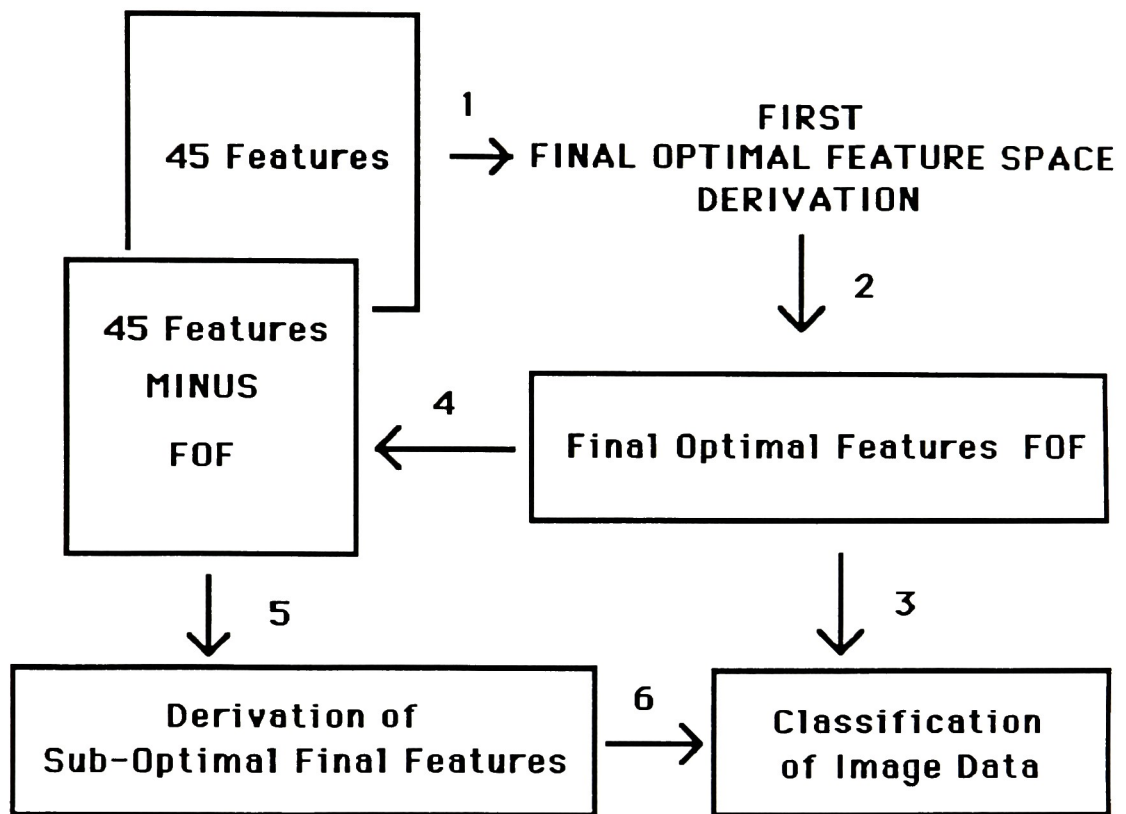
Step 1, 2 : Derivation of Final Optimal Feature Space for Analyses
162, 164, 322, 324

Step 3, 6 : Classification of Data in Final Optimal Feature Space

Step 4 : Removal of Final Optimal Features from 45 Feature Set

Step 5 : Derivation of Sub-Optimal Final Optimal Features for
Analyses 261, 461, 223, 423

Step 6 : Classification of Image Data in Sub-Optimal Final Optimal
Space for Analyses 261, 461, 223, 423

Figure   4.2.-11

Classification accuracies of identical training and test data for analyses (162, 261), (164, 461), (322, 223), and (324, 423) are shown in Figures 4.2-12 to 4.2-19 and in Appendices A-4 and A-5. Due to the "next-to optimal" characteristic of final optimal feature spaces for analyses 162, 164, 322, and 324, classification accuracies for "sub-optimal" final feature spaces for analyses 261, 461, 223, and, 423 would be comparable.
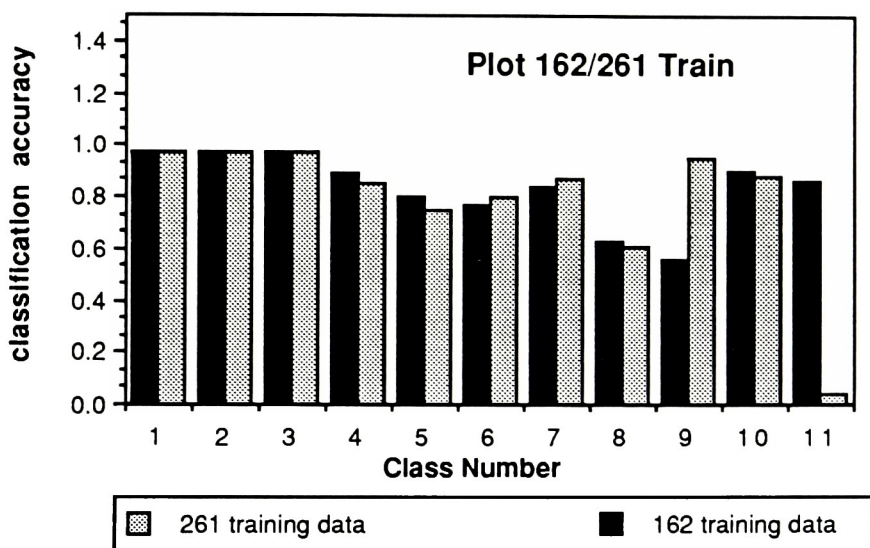
GCA 261 Training Data : 78.3%   **GCA 162 Training Data : 83.4%**
SUB-OPTIMAL   **OPTIMAL**

Figure   4.2-12

Plot 162/261 Test
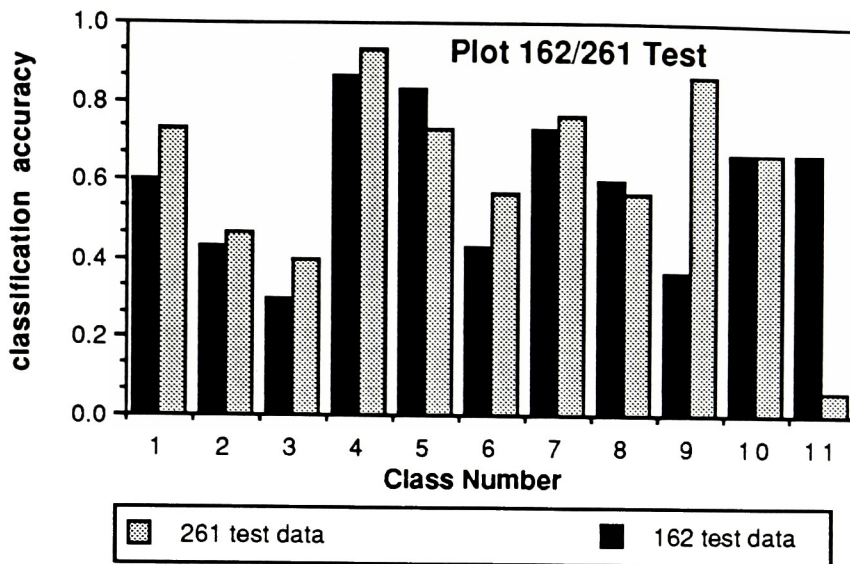
GCA 261 Test Data : 61.4% GCA 162 Test Data : 59.0%
SUB-OPTIMAL OPTIMAL

Figure   4.2-13

Plot 164/461 Train
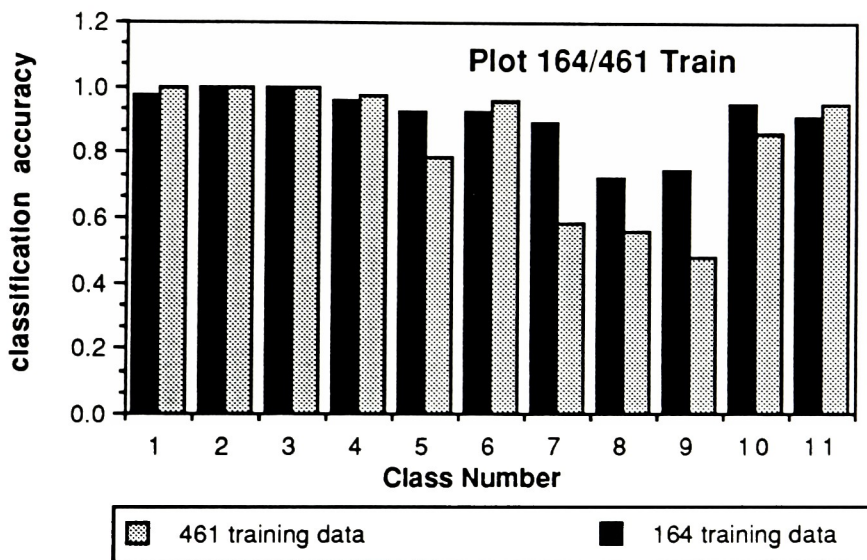
GCA 461 Training Data : 83.2% GCA 164 Training Data : 91.1%
SUB-OPTIMAL OPTIMAL

Figure   4.2-14
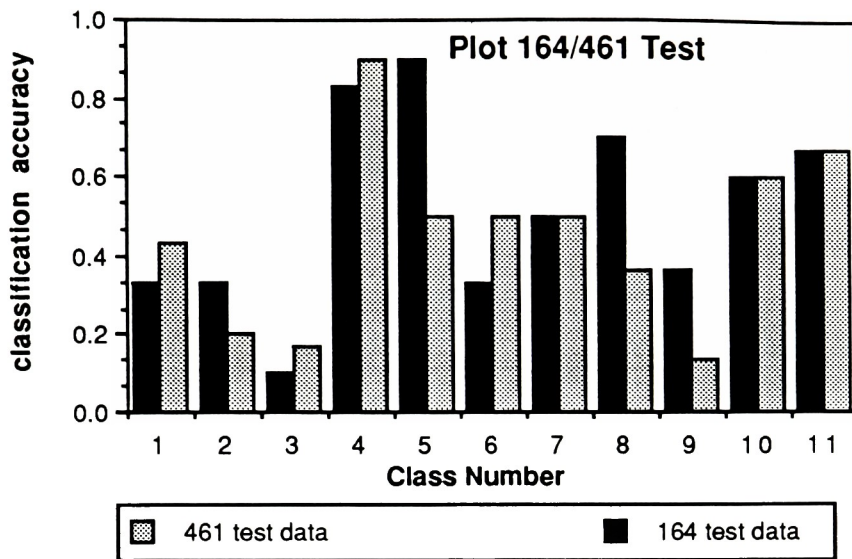
GCA 461 Test Data : 45.1%
SUB-OPTIMAL

GCA 164 Test Data : 51.5%
OPTIMAL

Figure   4.2-15
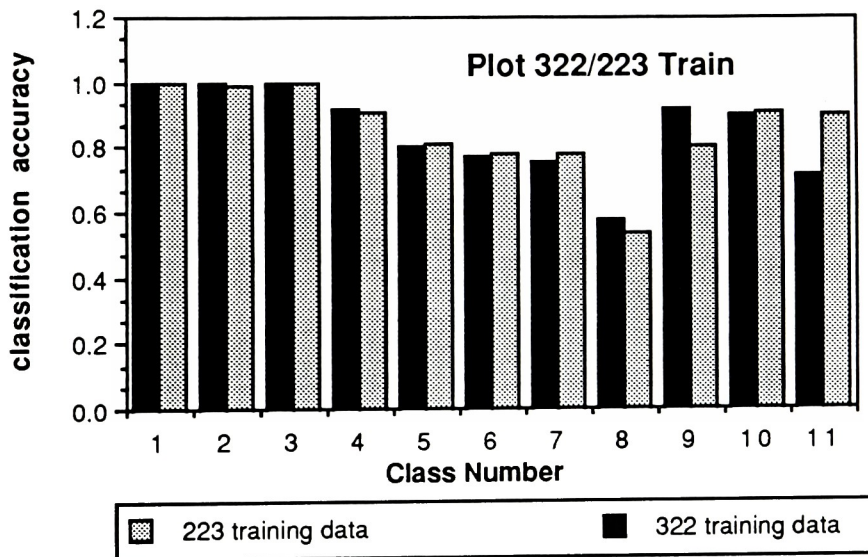
GCA 223 Training Data : 85.5%
SUB-OPTIMAL

GCA 322 Training Data : 85.1%
OPTIMAL

Figure   4.2-16

Plot 322/223 Test

Legend: 223 test data, 322 test data
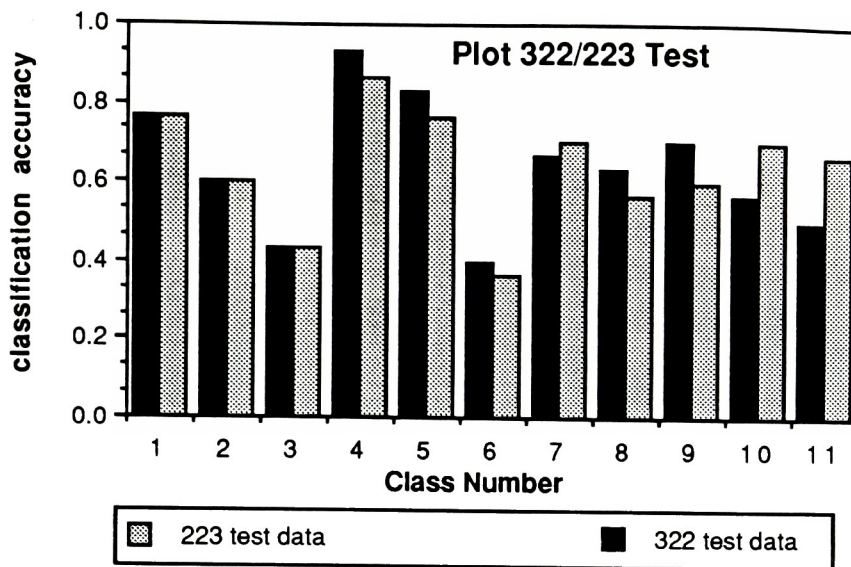
GCA 223 Test Data : 63.9%
SUB-OPTIMAL

GCA 322 Test Data : 63.9%
OPTIMAL

Figure   4.2-17

Plot 324/423 Train

Legend: 423 training data, 324 training data
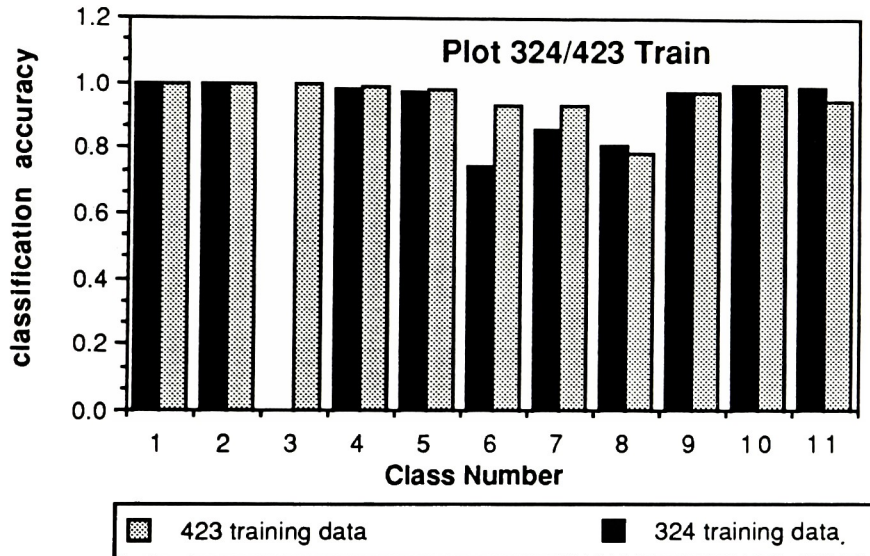
GCA 423 Training Data : 95.8%
SUB-OPTIMAL

GCA 324 Training Data : 85.0%
OPTIMAL

Figure   4.2-18
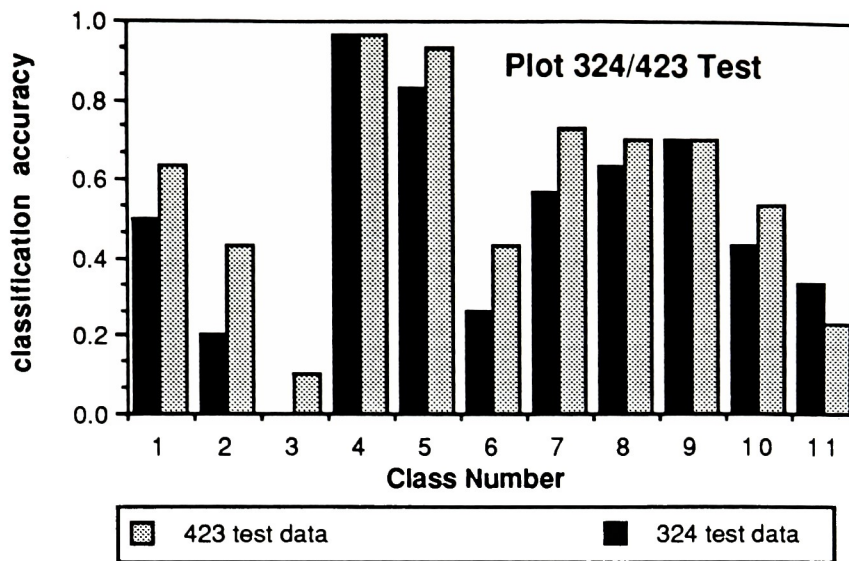
Figure 4.2-19

GCA 423 Test Data : 58.2%   GCA 324 Test Data : 49.4%
    SUB-OPTIMAL   OPTIMAL

4.2.3   Individual Class Covariance Matrices vs Class pooled Covariance Matrix

Use of individual class covariance matrices over class pooled covariance matrices for classification and final optimal feature selection was carried out.   Class pooled covariance matrices were derived for all analyses and image training data were classified in final optimal feature space.   Results are listed in Appendix A-6 and in plots of Figures 4.2-20 through 4.2-23 (in all plots CCM signifies Class Covariance Matrix).   Statistical hypothesis testing was performed for all analyses to prove inequality between individual

class covariance matrices.   Class pooled covariance matrices could
be used if test results were false.

Plot Roch162 (Pool vs Indiv.)

Legend: 162 training data/pool   162 training data/indiv.
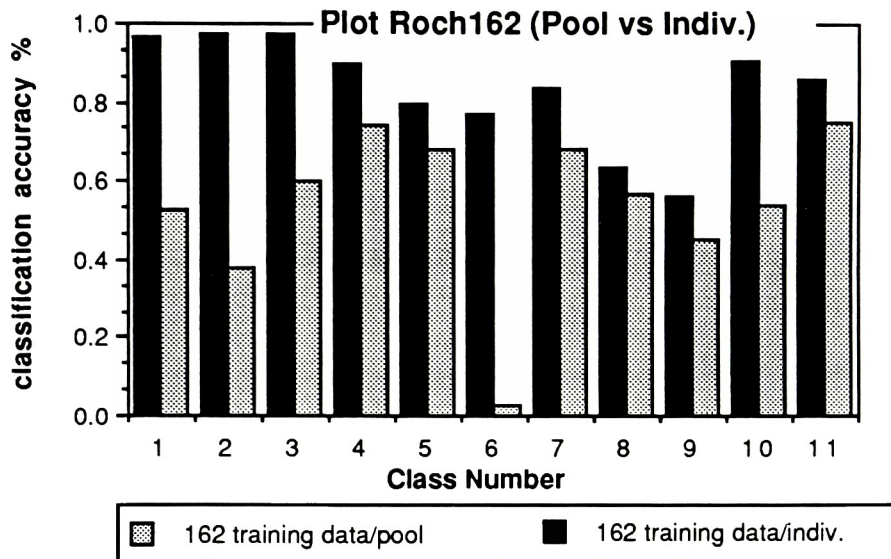
GCA 162 Individual CCM : 83.4%   GCA 162 Pooled CCM : 54.6%

Figure   4.2-20

Plot Roch164 (Pool vs Indiv.)

Legend: 164 training data/pool | 164 training data/indiv.
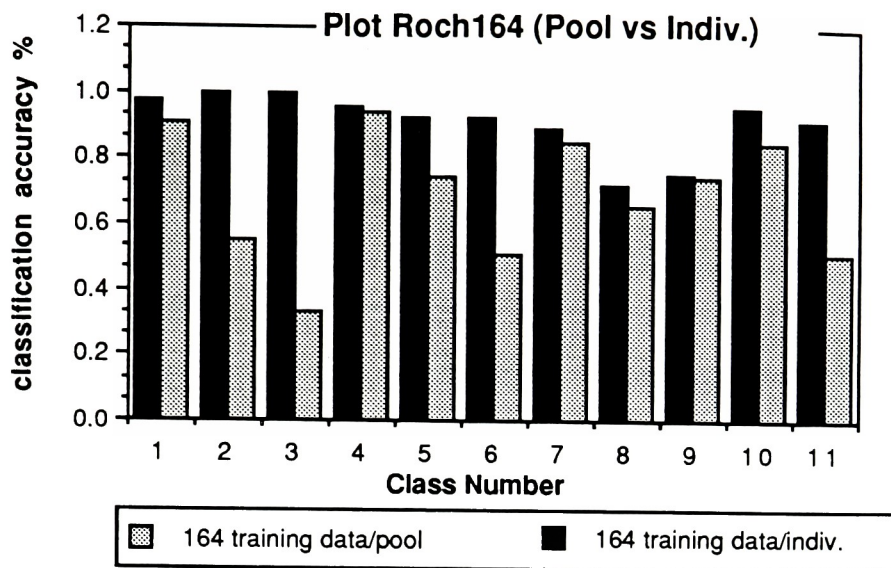
GCA 164 Individual CCM : 91.1%   GCA 164 Pooled CCM : 69.5%

Figure   4.2-21

Plot Roch322 (Pool vs Ind)

Legend: 322 training data/pool | 322 training data/indiv.
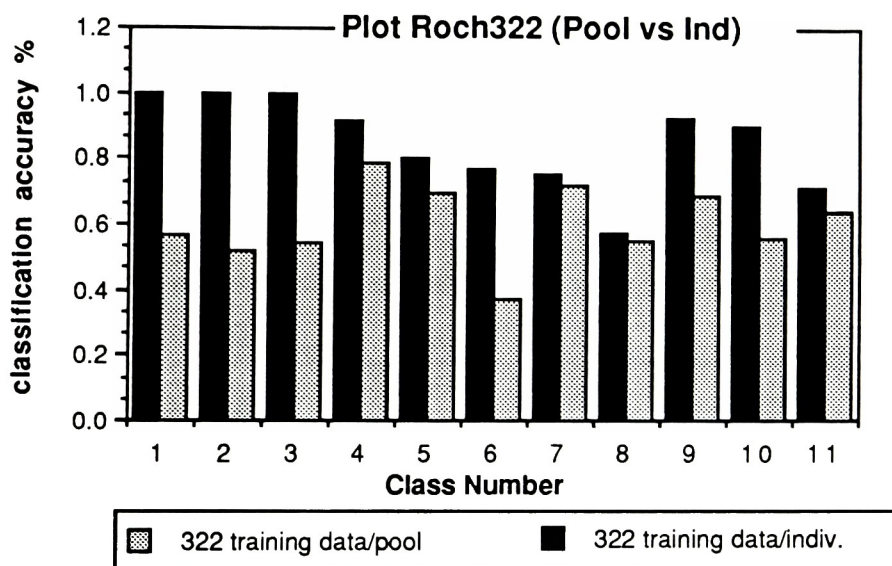
GCA 322 Individual CCM : 85.1%   GCA 322 Pooled CCM: 60.8%

Figure   4.2-22

Plot Roch324 (Pool vs Ind)

Legend: 324 training data/pool ◻   324 training data/indiv. ■
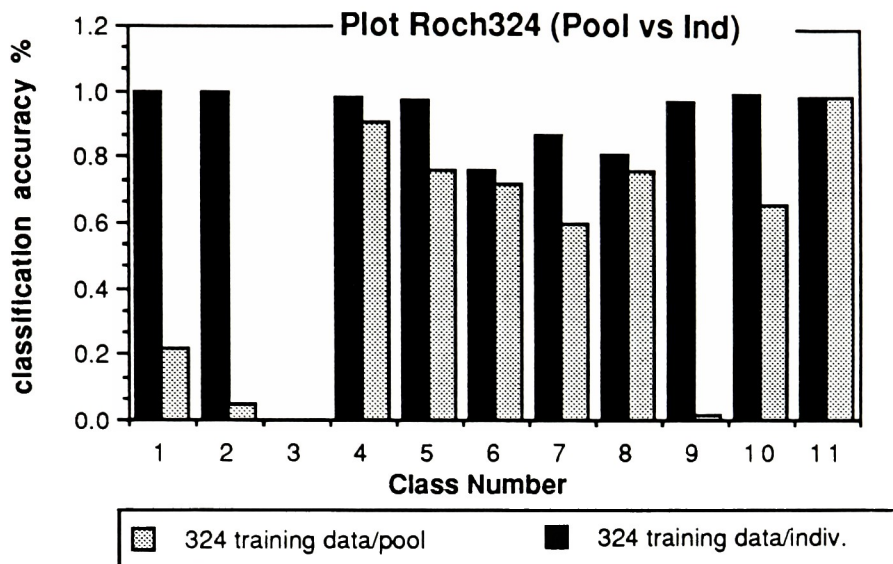
GCA 324 Individual CCM : 85.0%   GCA 324 Pooled CCM: 52.0%

Figure   4.2-23

4.2.4   Validation of Final Optimal Feature Space Construction Process

Validation of the process by which final optimal feature spaces were constructed was carried out for analyses 162, 164, and 322. For analysis 162, the initial optimal feature space consisting of 11 textural features was divided into two subsets containing 5 features apiece. Subset #1 contained optimal features derived from discriminant analysis while subset #2 contained an equal number of the remaining initial optimal features. Similar processing was carried out for analyses 164 and 322. Two subsets, containing 8 features apiece were established for analysis 164 and 6 features

90

apiece for analysis 322. Individual feature subsets for analyses 162, 164, and 322 are listed in Appendix A-7.

For each subset, the initial training and test data files were restructured to contain their individual features. Class statistics were derived to train maximum likelihood classifiers and image training and test data were classified. Figure 4.2-24 illustrates this process, where the abbreviation FOFS signifies Final Optimal Feature Space. Classification results are listed in Figures 4.2-25 through 4.2-30 and in Appendix A-8.

VALIDATION OF FINAL OPTIMAL FEATURE SPACE CONSTRUCTION PROCESS

Figure   4.2-24

Plot Roch162 Subsets (training data)
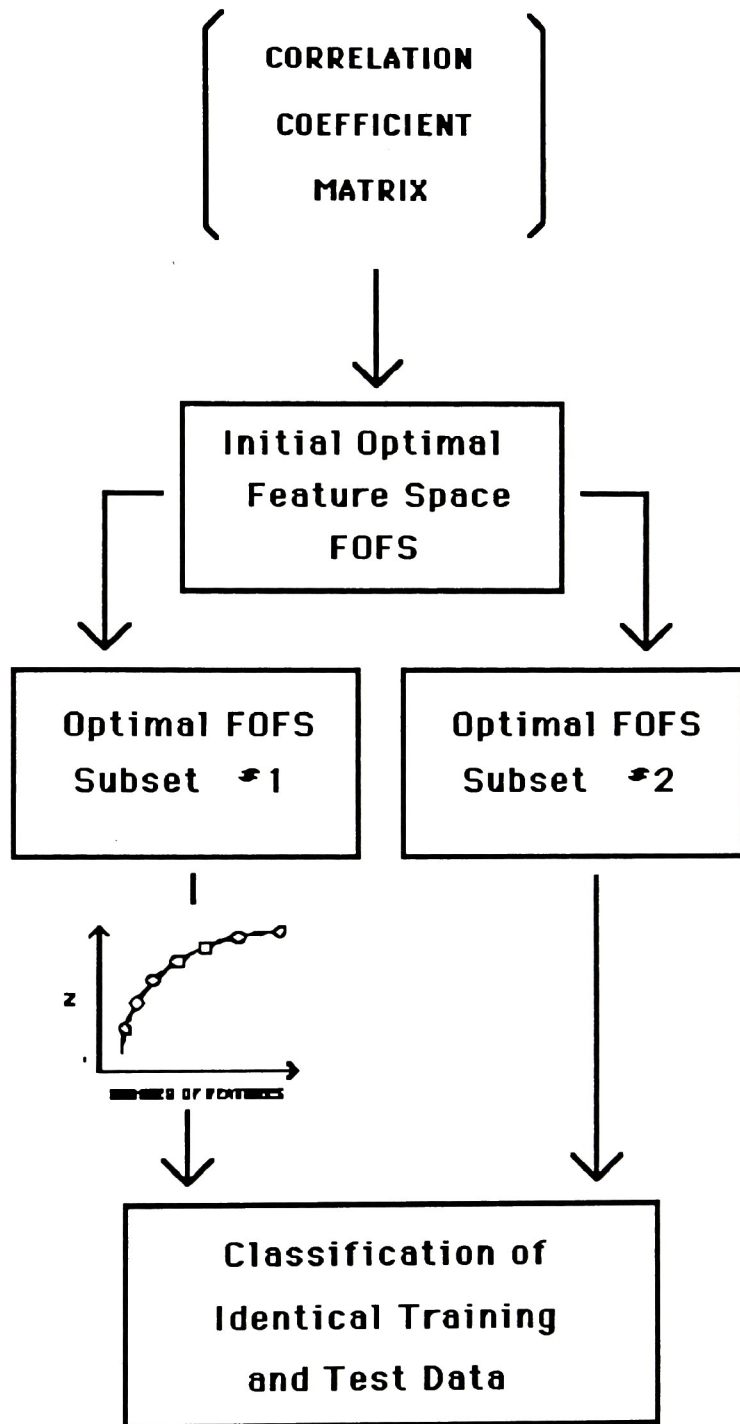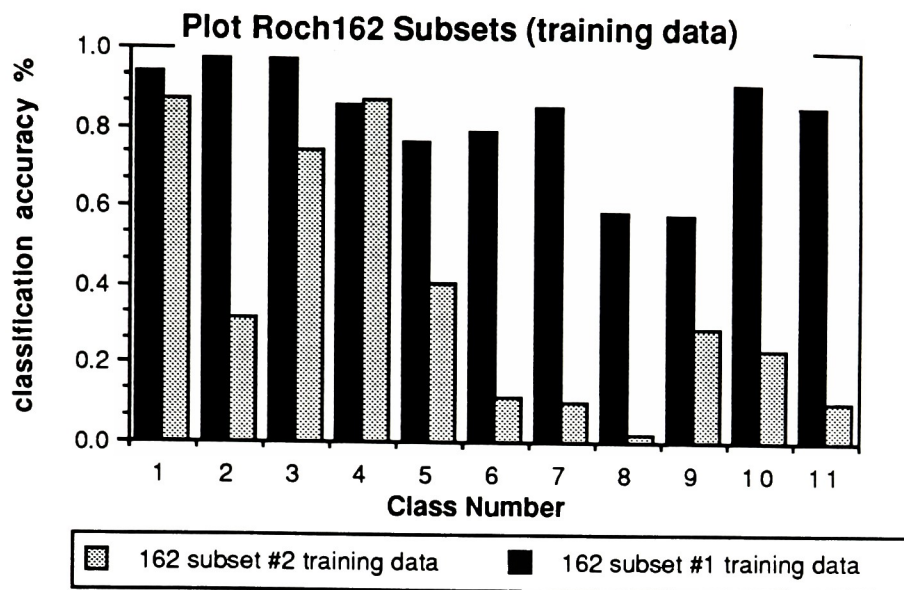
GCA 162 Subset #1 :   82.6%   GCA 162 Subset #2 :   37.0%

Figure   4.2-25

Plot Roch162 Subsets (test data)
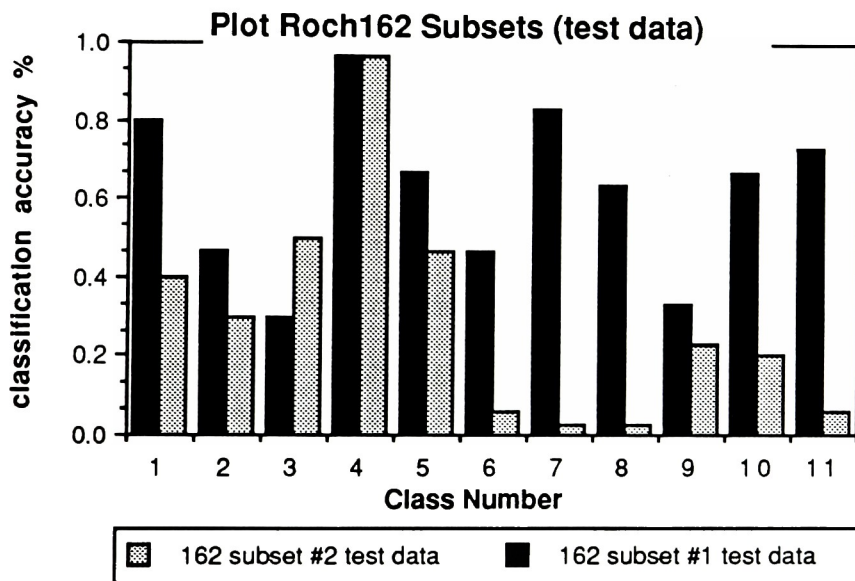
GCA162 Subset #1 :   62.4%   GCA 162 Subset #2 :   29.5%

Figure   4.2-26

Plot Roch164 Subsets (training data)
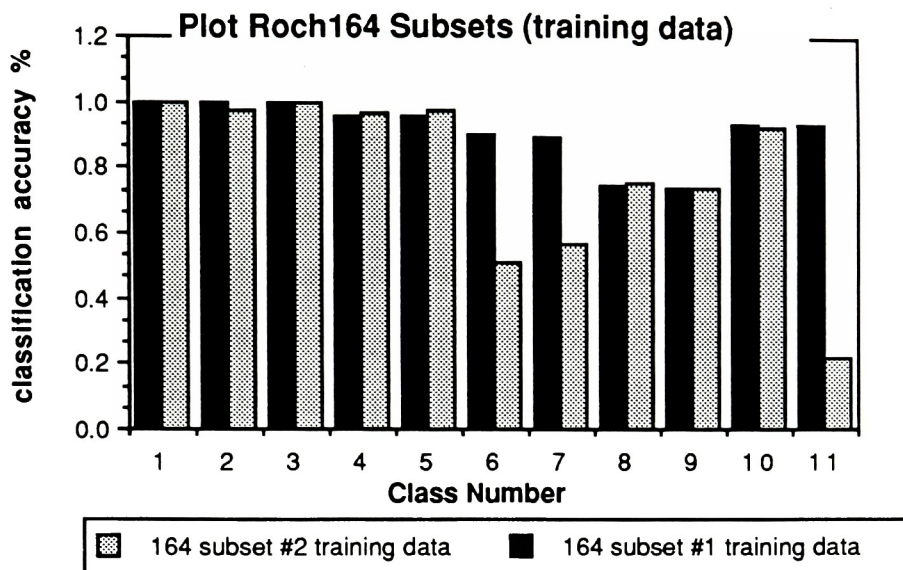
GCA 164 Subset #1 : 91.5%   GCA 164 Subset #2 : 78.5%

Figure   4.2-27

Plot Roch164 Subsets (test data)
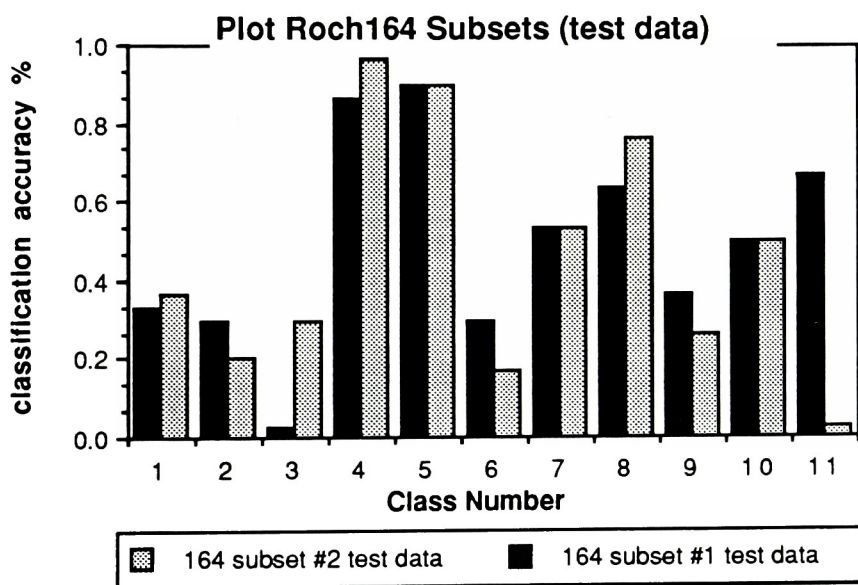
GCA 164 Subset #1 : 49.3%   GCA 164 Subset #2 : 45.4%

Figure   4.2-28

Plot Roch322 Subsets (training data)
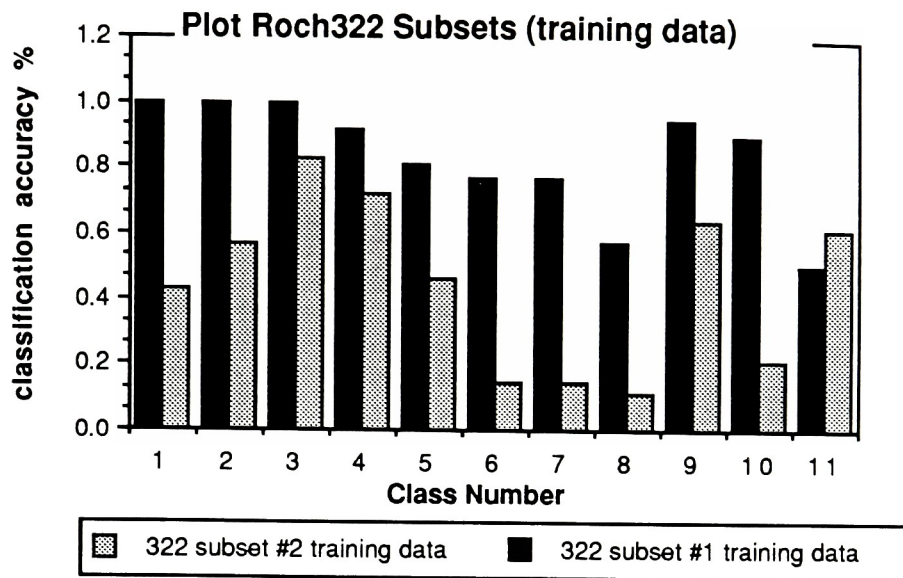
GCA 322 Subset #1 : 83.6%   GCA 322 Subset #2 : 44.4%

Figure   4.2-29

Plot Roch322 Subsets (test data)
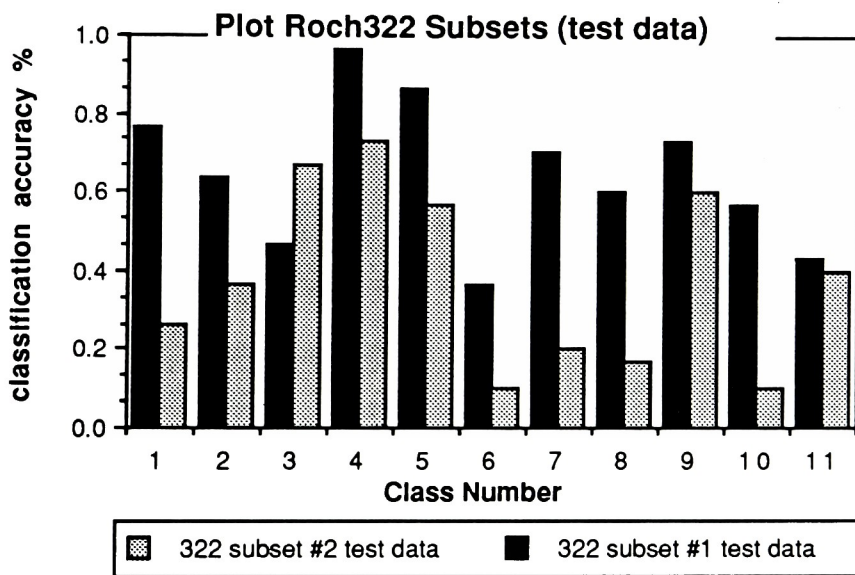
GCA 322 Subset #1 : 64.5%   GCA 322 Subset #2 : 37.8%

Figure   4.2-30

4.3   Processing of Image Field

Seven ground cover classes were identified within image **Field** entitled : 1) light grey desert, 2) dark grey shrubs, 3) light grey shrubs, 4) light grey stone, 5) light grey grass, 6) dark grey desert, and 7) interspersed dark grey bushes and surrounding desert.   All classes displayed overall uniform texture with exception to class 7, which included a mix of bush and desert textures.   Image Field is shown in Figure 4.3-1.
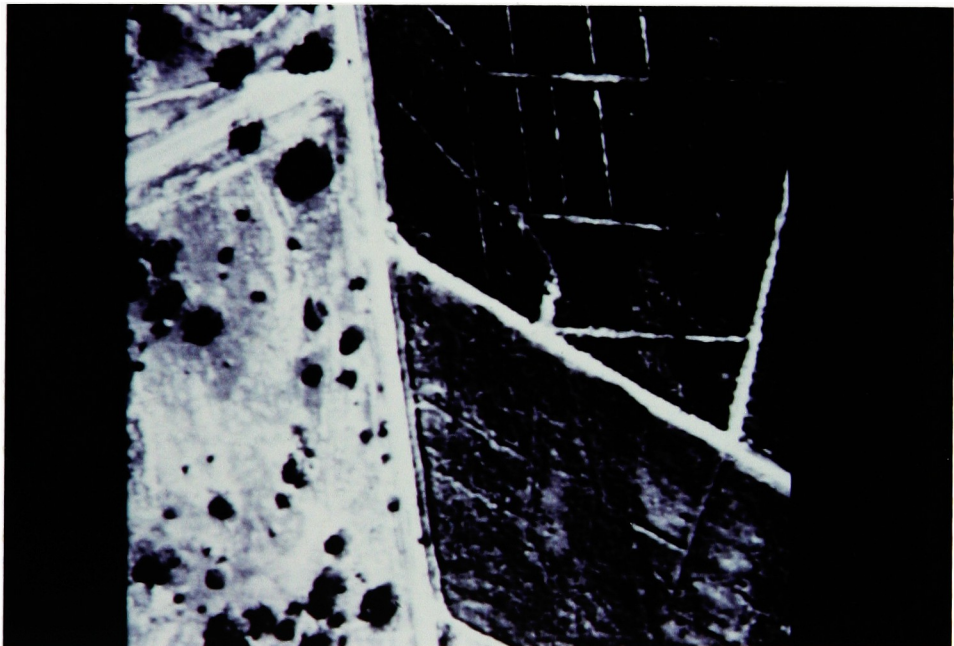
Figure   4.3-1

Similar processing was carried out on image Field as was performed on image Urban. Fourteen image class windows were selected across the seven classes to encompass training pixels to train classifiers. For class 7, entitled bushes/desert, large image class windows were designed to included both desert and tree type pixels. This allowed for use of large image sampling windows over all classes.

4.3.1 Analyses 17, 21, 25, 3225, 21*3, 25*5, 21*r

Image **Field** was processed in 7 ways entitled : 17, 21, 25, 3225, 21*3, 25*5, and 21*r as listed in Table 4.3-1. Analyses (17, 21, 25) displayed the effects of increased image sampling window size and constant number of grey tone quantization levels on classification accuracy. Analyses (25, 3225) illustrated similar effects from constant image sampling window size and increased number of grey tone quantization levels. Analyses (21, 21*3) and (25, 25*5) displayed the effects of increased image magnification by enlargement of the interpixel sampling interval to every third pixel for analysis 21*3 and to every fifth pixel for analysis 21*5. Analyses (21, 21*r) displayed the effects of a 30 degree clockwise rotation.

For all analyses, final optimal feature spaces were established and image test and training data were classified (note : the final optimal feature space for analysis 21*r consisted of that from analysis 21).

TABLE 4.3-1

Analysis Parameter Settings for Analyses 17, 21, 25, 21*3, 25*5, 3225, 21*r over Image Field

	Analysis			
	17	**21**	**25**	**21*3**
Parameters				
Quantization Levels	16	16	16	16
Sampling Window Size	17*17	21*21	25*25	21*21
Interpixel Sampling	1	1	1	3
Degrees of Rotation	0	0	0	0
	25*5	**3225**	**21*r**	
Quantization Levels	16	32	16	
Sampling Window Size	25*25	25*25	21*21	
Interpixel Sampling	5	1	1	
Degrees of Rotation	0	0	30 cw	

note : cw = clockwise rotation

With exception to analysis 21*r, the logic of Appendix E was followed for all analyses. Forty-six textural features were derived for each pixel contained within the boundaries of all image class windows. Textural features from all processes for image Field are listed in Appendix A-1. The distribution of training pixels per class for each analysis is listed in Table 4.3-2. Limited numbers of training pixels for several classes was due to small class size compared to the sizes of image class and sampling windows.

TABLE 4.3-2

Number of Training Pixels per Class for Analyses 17, 21, 25, 21*3, 25*5, 3225 for Image Field

Analysis :	17	21/21*3	25/25*5/3225
Image Class			
1) light grey desert	1152	800	512
2) dark grey shrubs	1176	820	528
3) light grey shrubs	1152	800	512
4) light grey stone	576	400	256
5) light grey grass	392	200	72
6) dark grey desert	322	150	42
7) mix bushes/desert	837	485	229

For all analyses, class individual covariance matrices and a class pooled covariance matrix of dimension 46*46 were derived, and feature correlation coefficient matrices were extracted. In accordance with the correlation coefficient matrix feature selection technique of Section 2.4.1, a threshold value of 60% was established above and beyond which all correlation coefficients sharing a common second dimension were grouped. Fifteen initial optimal features were derived for analysis 17, 16 features for analysis 21, 14 features for analysis 25, 18 features for analysis 21*3, 16 features for analysis 21*5, and 15 features for analysis 3225. For analysis 21, initial optimal features were measured over all image pixels and textural feature images were derived. Initial optimal feature spaces are listed in Appendix B-1, and textural feature images for analysis 21 are displayed in Appendix C-6.

The training data files were restructured to contain initial optimal features. These data files were processed to derive class mean vectors, individual class covariance matrices, optimal feature combination subsets and associated maximum divergence statistics Z. Final optimal features were derived from analysis of maximum divergence statistics in accordance with the feature selection methods of Section 2.4.1. Figures 4.3.-2 through 4.3-7 show the dimensions of final optimal feature spaces in terms of Maximum Divergence vs Number of Features in Final Optimal Feature Space. Circled data points represent these dimensions. Final optimal feature spaces are listed in Appendix B-1. Figure 4.3-8 lists the

100

dimensions of initial and final optimal feature spaces. Best feature combinations for each analysis are listed in Appendix B-6.
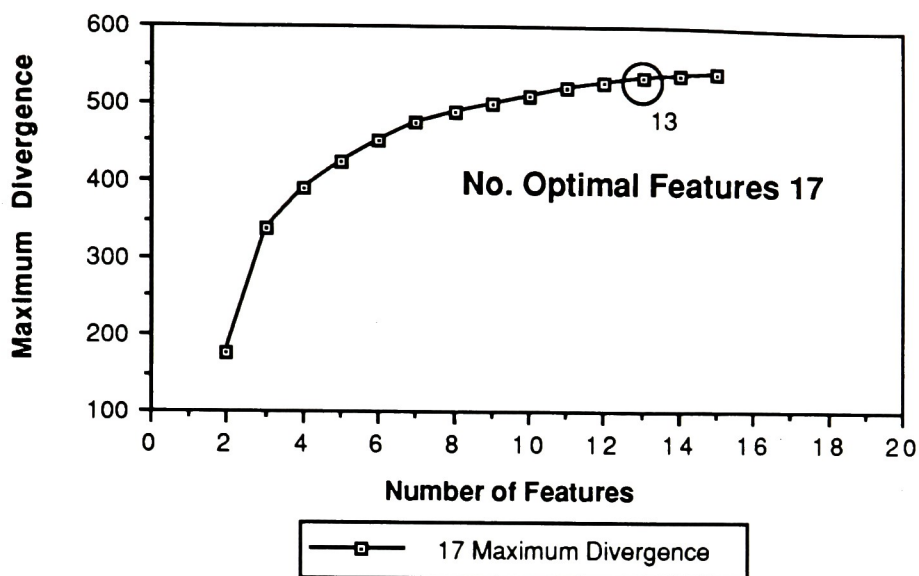
Figure 4.3-2
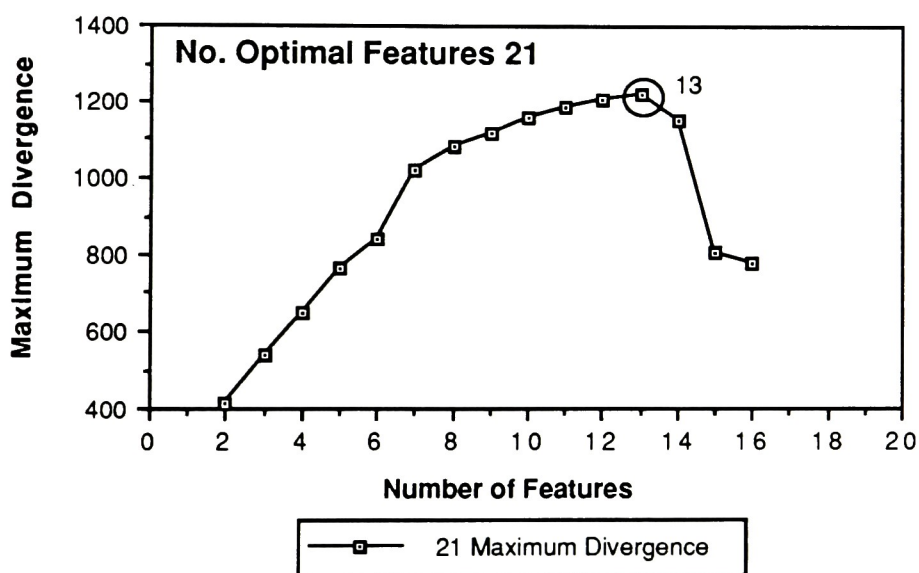
Figure 4.3-3
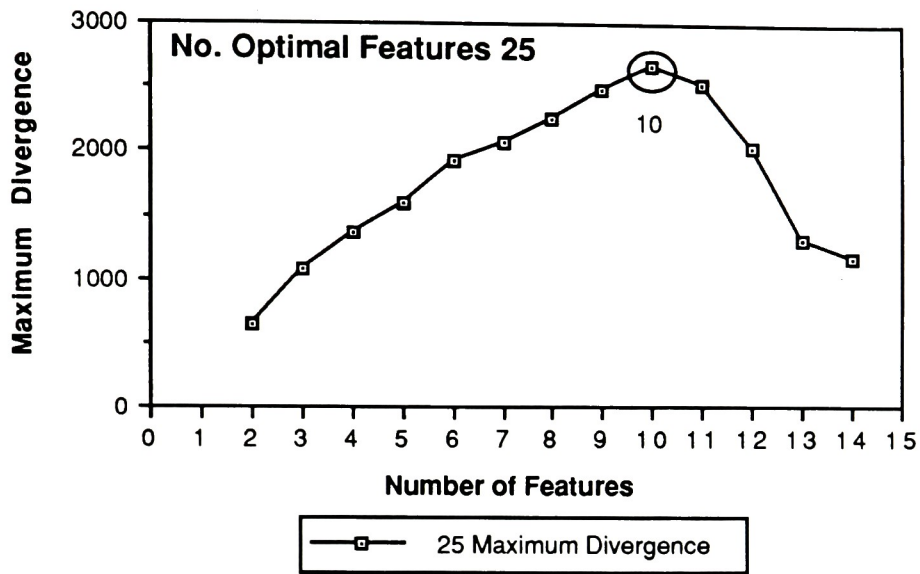
Figure 4.3-4
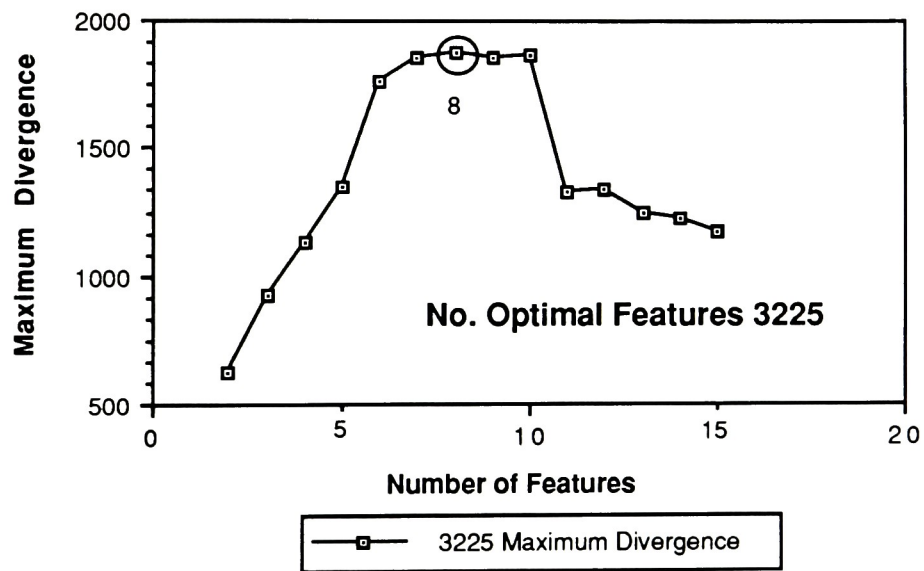
Figure 4.3-5
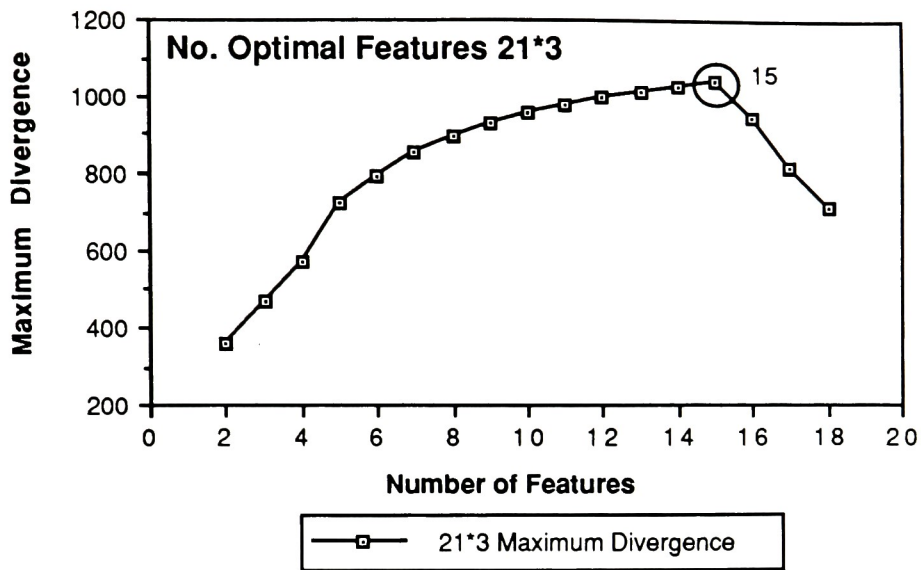
Figure 4.3-6
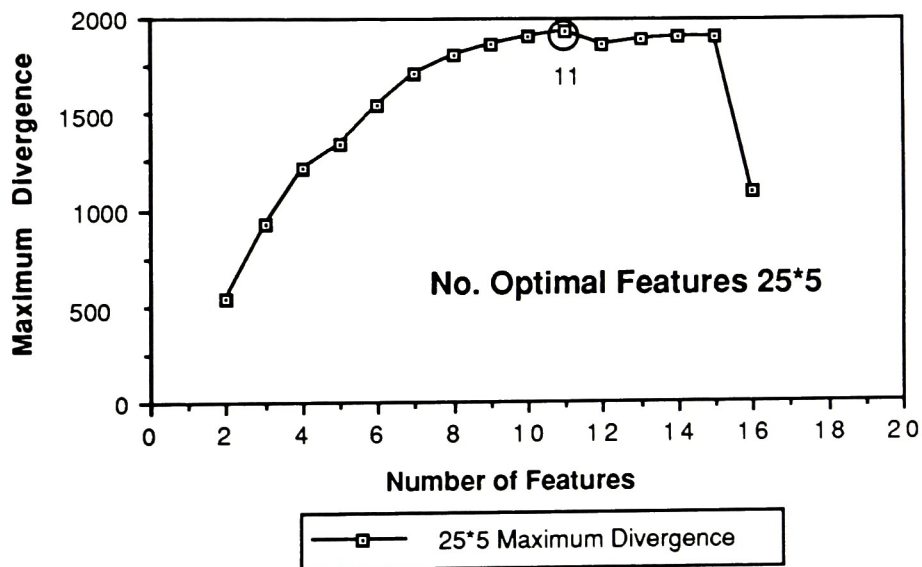
Figure 4.3-7

INITIAL AND FINAL OPTIMAL FEATURE SPACES

17	21	25
46 feat	46 feat	46 feat

21*3	25*5	3225
46 feat	46 feat	46 feat

INITIAL OPTIMAL FEATURE SPACE

15 f   18 f   16 f   16 f   14 f   15 features

FINAL OPTIMAL FEATURE SPACE

13f   15f   13f   11f   10f   8features
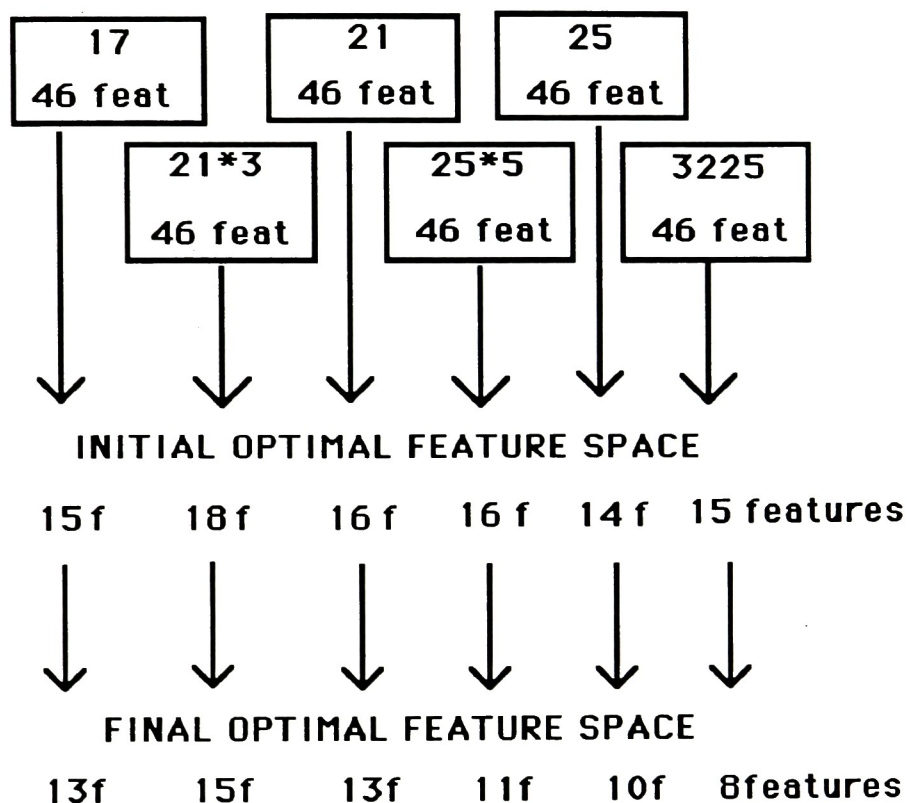
Figure 4.3-8

For each analysis, the training data was restructured to include final optimal features. Class statistics were derived and image training and test data were classified. Test data consisted of thirty pixels per class, selected at random by the analyst, across the face of each class. Results of the classification processes in terms of

in Appendix B-2 and displayed in Figures 4.3-9 through 4.3-14. The term "Field" in all plots relates to analyses over image Field.
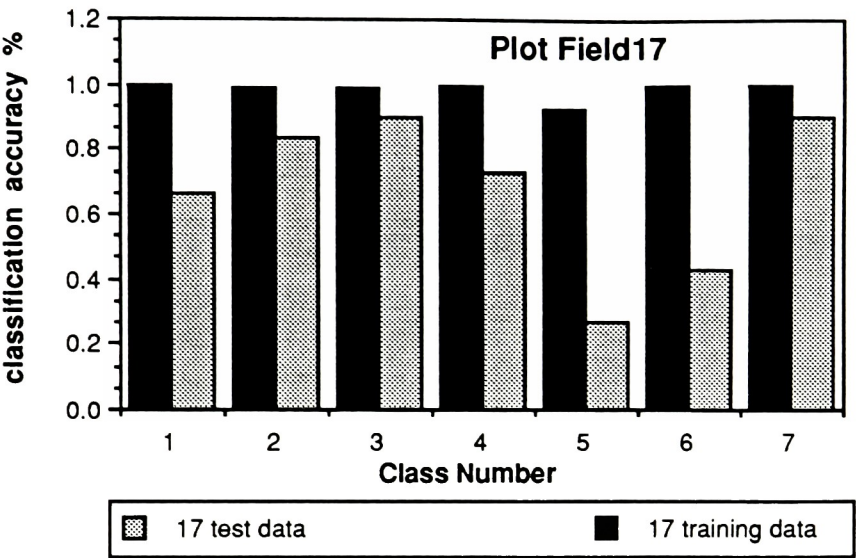
GCA 17 Training Data : 98.7%   GCA 17 Test Data : 67.6%

Figure 4.3-9

Plot Field21
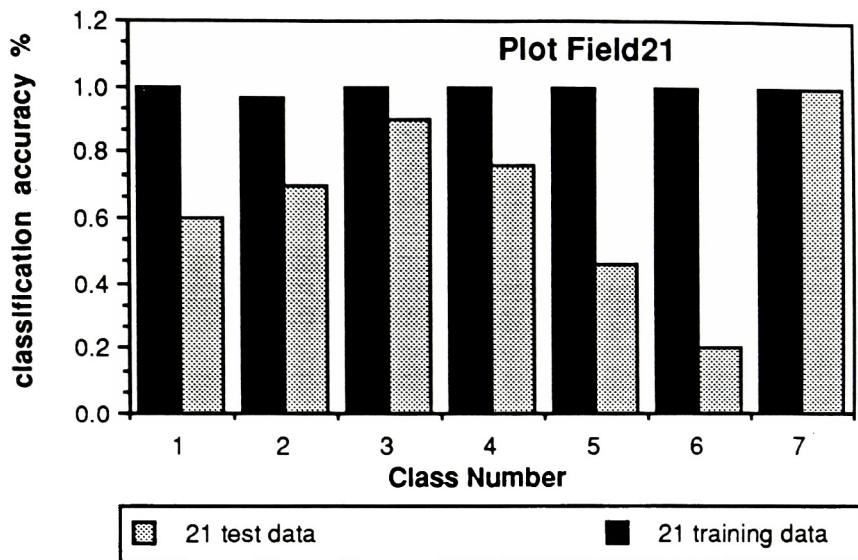
GCA 21 Training Data : 99.4%   GCA 21 Test Data : 66.2%

Figure   4.3-10

Plot Field25
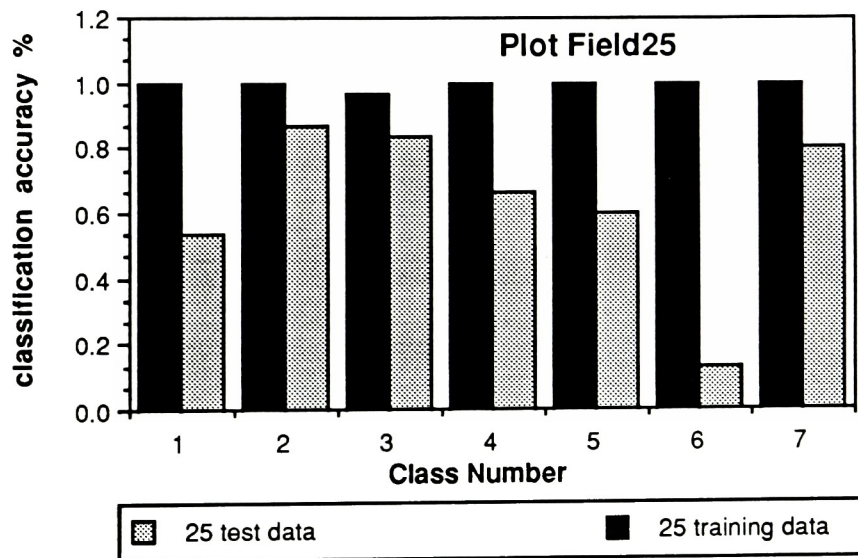
GCA 25 Training Data : 99.5%   GCA 25 Test Data : 63.3%

Figure   4.3-11

Plot Field21*3
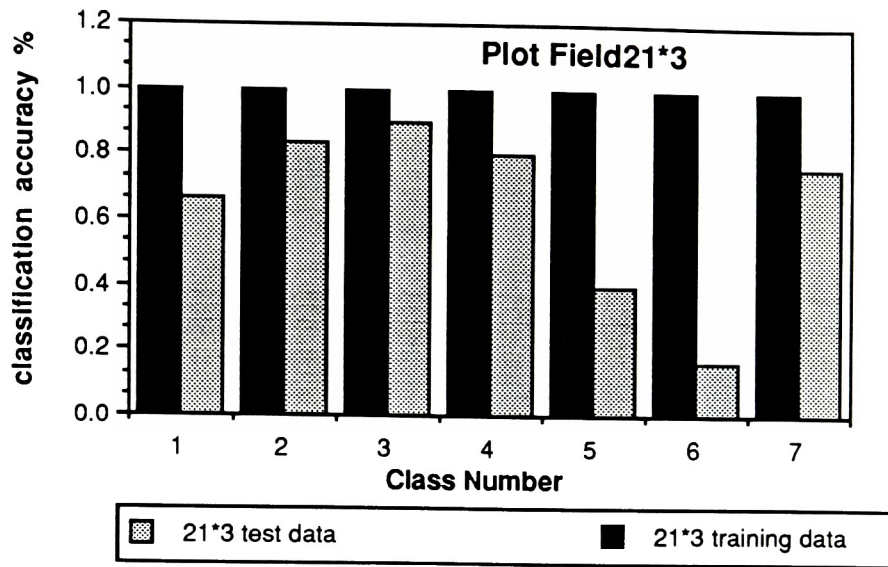
GCA 21*3 Training Data : 99.9%   GCA 21*3 Test Data : 64.7%

Figure   4.3-12

Plot Field25*5
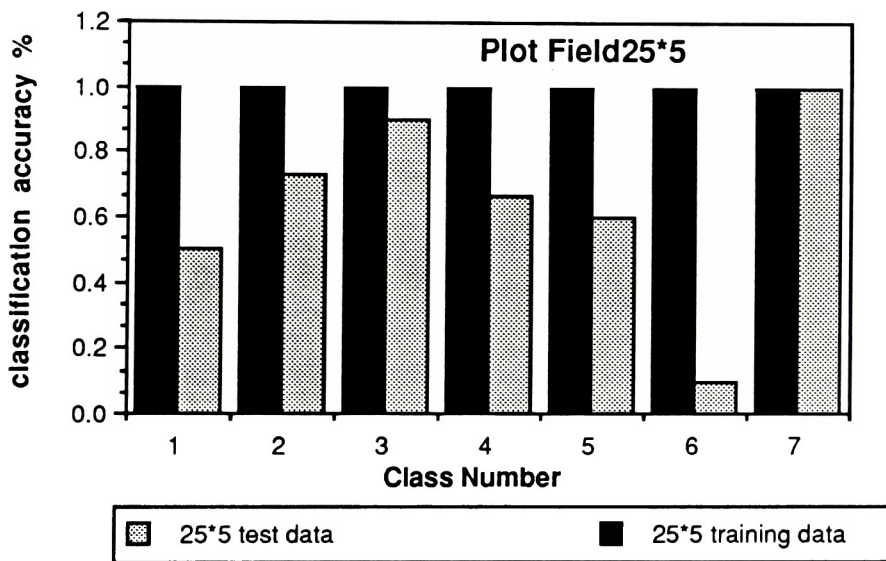
GCA 25*5 Training Data : 99.9%   GCA 25*5 Test Data : 64.3%

Figure   4.3-13

108

Plot Field3225

Legend: ▨ 3225 test data   ■ 3225 training data
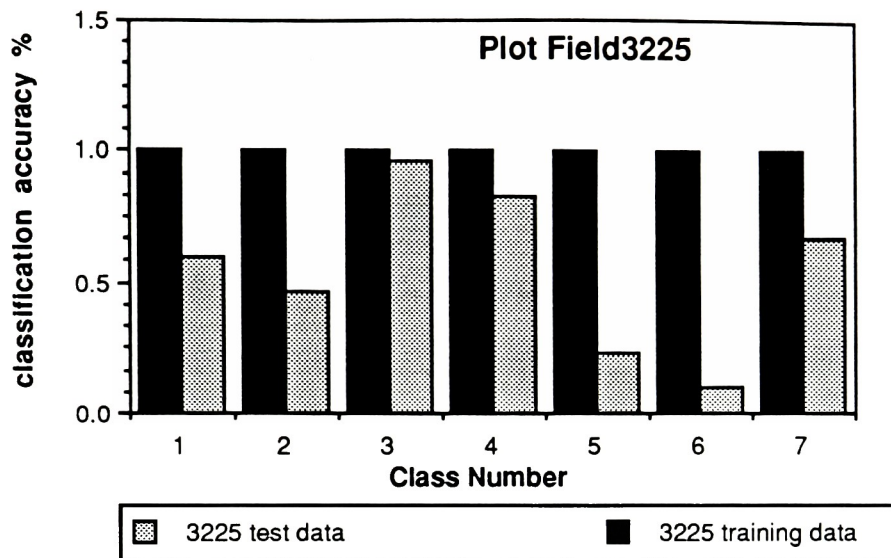
GCA 3225 Training Data :100%   GCA 3225 Test Data : 55.2%

Figure   4.3-14

4.3.2   Individual Class Covariance Matrices vs Pooled Class Covariance Matrix

Classification and feature selection exercises, using class pooled covariance matrices, were not carried out as to ensure maximum accuracy for all analyses.   By statistical hypothesis testing, it was determined that individual class covariances were unequal which excluded class pooled covariance matrices from consideration. Section 4.2.3 provides further argument.

4.3.3   Analysis 21*r

Thirty test pixels were selected at random by the analyst within each of the seven classes of the rotated image. Each test pixel was characterized by final optimal features of analysis 21, and was derived from image sampling windows of dimension 21*21 which were oriented at 0 degrees with respect to the image. Classification results of image test data from analyses 21*r and 21 are listed in Appendix B-2 and are displayed in Figure 4.3-15.
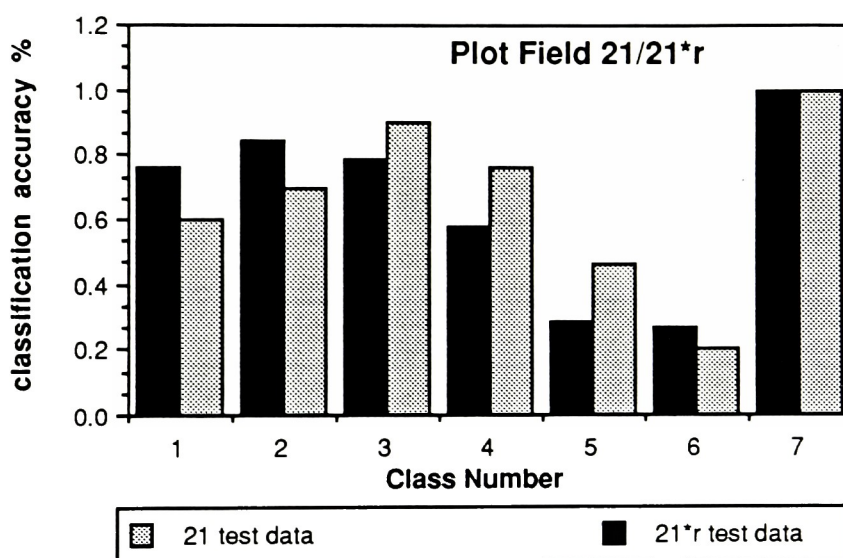
GCA 21 Test Data : 66.2%   GCA 21*r Test Data : 64.7%

Figure   4.3-15

4.3.4 Validation of Final Optimal Feature Space Construction Process

The optimality of final optimal feature spaces was tested for analyses 21 and 25. Two subsets of features were formed for each analysis from initial optimal feature spaces. Subsets #1 and #2 contained 8 features apiece for analysis 21, and 6 features apiece for analysis 25. Subsets #1 contained optimal features derived from discriminant analysis while subsets #2 contained an equal number of the remaining features from the initial optimal feature space. Individual feature subsets for analyses 21 and 25 are listed in Appendix B-3.

Training and test data files were restructured for analyses 21 and 25 to contain the features of their respective subsets. Individual class statistics were derived and image training and test data were classified. Results from these classification exercises are listed in Figures 4.3-16 through 4.3-19 and in Appendix B-4. Figure 4.2-24 illustrates the process used for validation of the optimality of final optimal feature spaces.

Plot Field21 Subsets (training data)
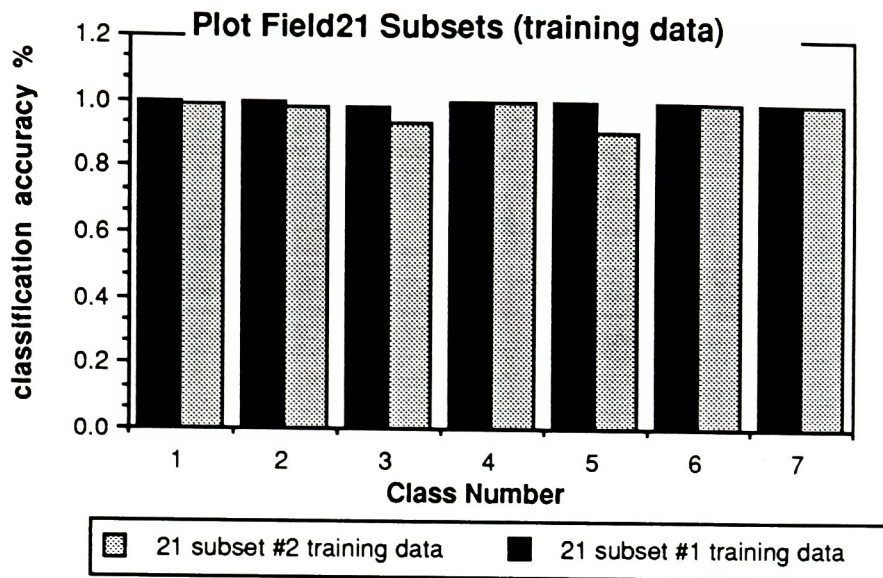
GCA 21 Subset #1 : 99.7%   GCA 21 Subset #2 : 97.2%

Figure   4.3-16

Plot Field21 Subsets (test data)
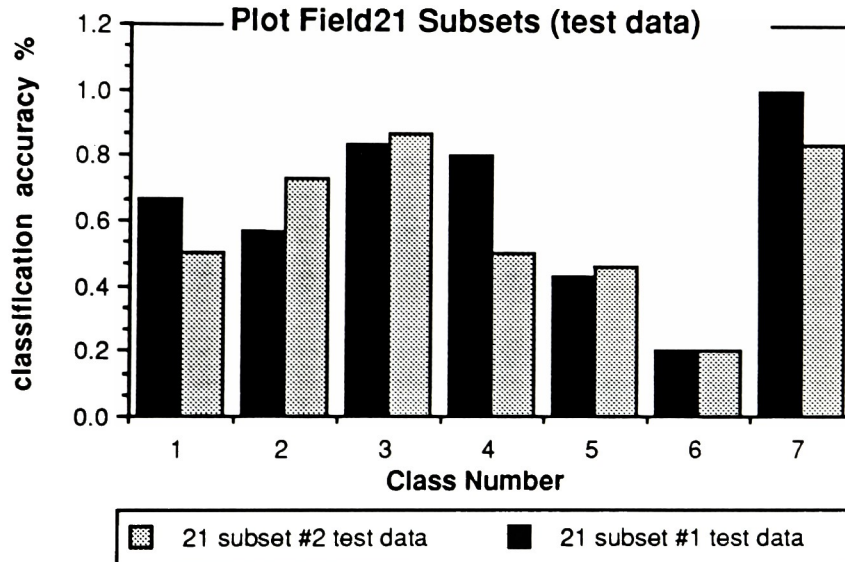
GCA 21 Subset #1 : 64.3%   GCA 21 Subset #2 : 58.5%

Figure   4.3-17

Plot Field25 Subsets (training data)
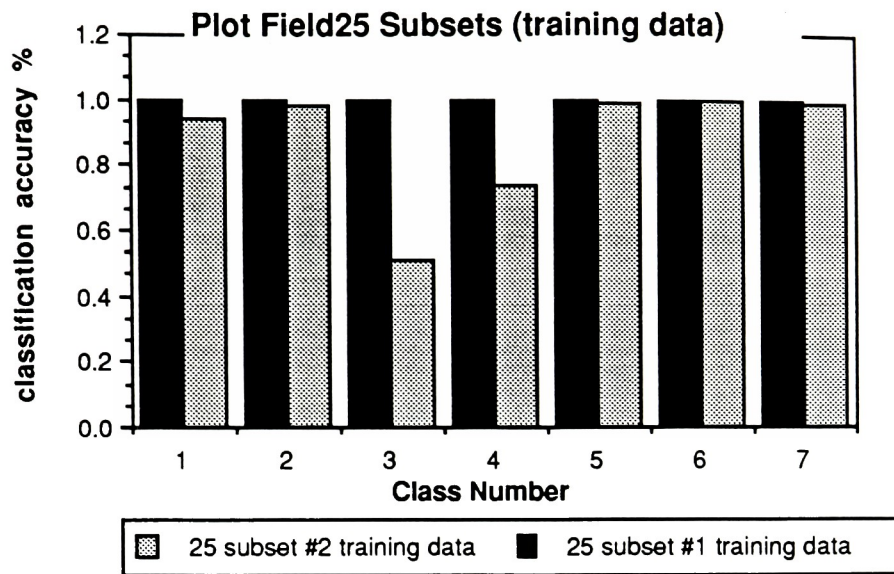
GCA 25 Subset #1 : 99.1%   GCA 25 Subset #2 :   87.8%

Figure   4.3-18

Plot Field25 Subsets (test data)
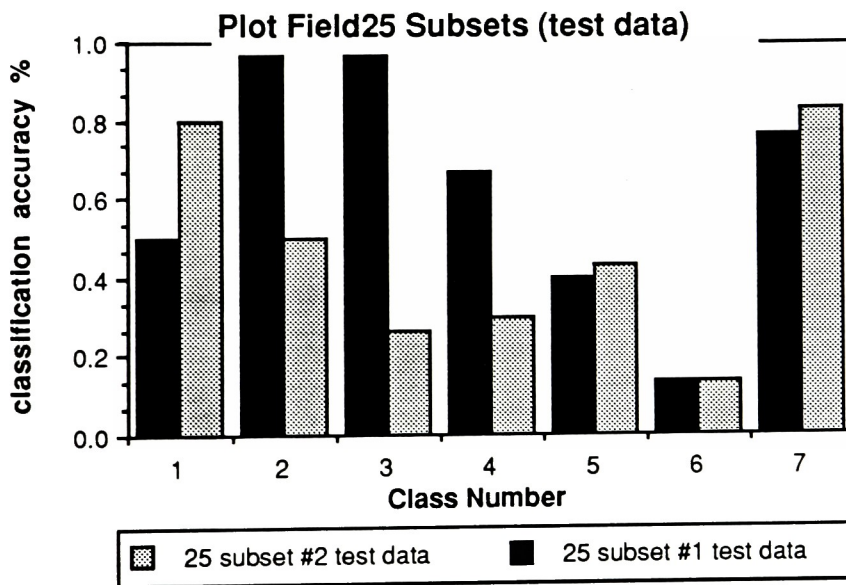
GCA 25 Subset #1 :   62.8%   GCA 25 Subset #2 :   46.6%

Figure   4.3-19

4.3.5 Computer Selected Test Data and Full Scene Gaussian Maximum Likelihood Classification

Objective and random selection of test data was carried out for analyses 17 and 21 using the ERDAS software library. This approach allowed for comparison of global classification accuracies of randomly selected test data sets; one selected by the analyst and the other by the computer. The results illustrate the effects of mixed pixels on classification accuracy. Mixed pixels are difficult to identify due to their location within the image (i.e. along class boundaries). Twenty test pixels per class were selected at random by the computer and were classified in final optimal feature space. Classification results of computer (denoted by the title ERDAS) and analyst selected test data for analyses 17 and 21 are shown in Figures 4.3-20 and 4.3-21 and are listed in Appendix B-5.
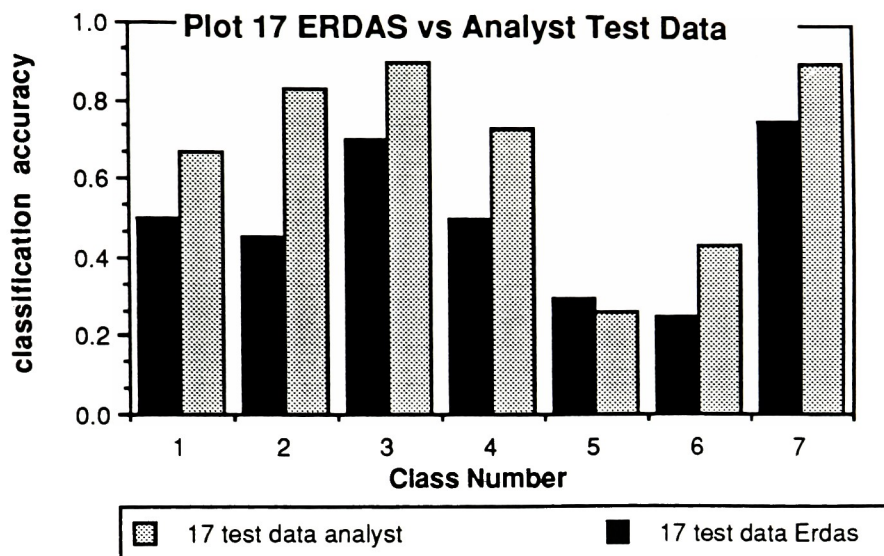
GCA Test Data Analyst : 67.6%   GCA Test Data ERDAS : 49.3%

Figure   4.3-20
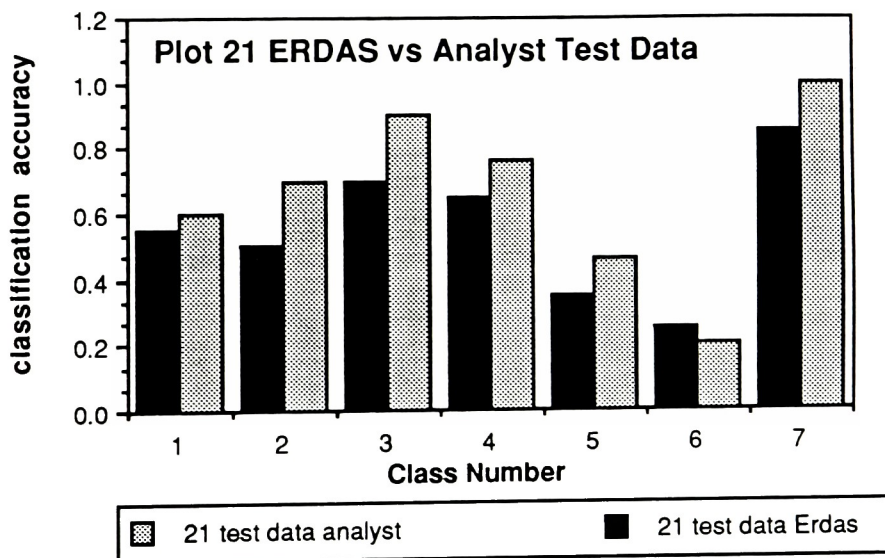
GCA Test Data Analyst : 66.2%   GCA Test Data ERDAS : 55.0%

Figure   4.3-21

Maximum Likelihood classification was carried out for analysis 21 using a computer developed classifier. Textural feature images of final optimal features entitled angsecaver, angsecrang, corraver, corrang, varrang, moyenne, and ecart were used to train the classifier. Computer selected test data was classified by the computer and identified by the analyst for determination of classification accuracy. Results of classification are illustrated in Figure 4.3-22 and in Appendix B-5 for 7 textural feature image training set. The abbreviation FOFSC stands for classification in final optimal feature space of computer selected image data, and ERDAS stands for classification performed by the computer. The classified image is displayed in Figure 4.3-23.
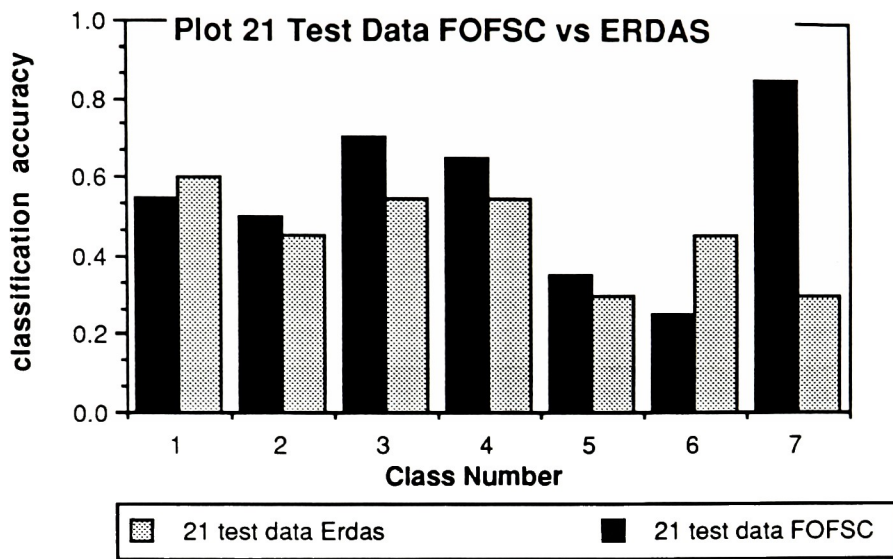
GCA Test Data ERDAS : 45.7%   GCA Test Data FOFSC : 55.0%

Figure   4.3-22
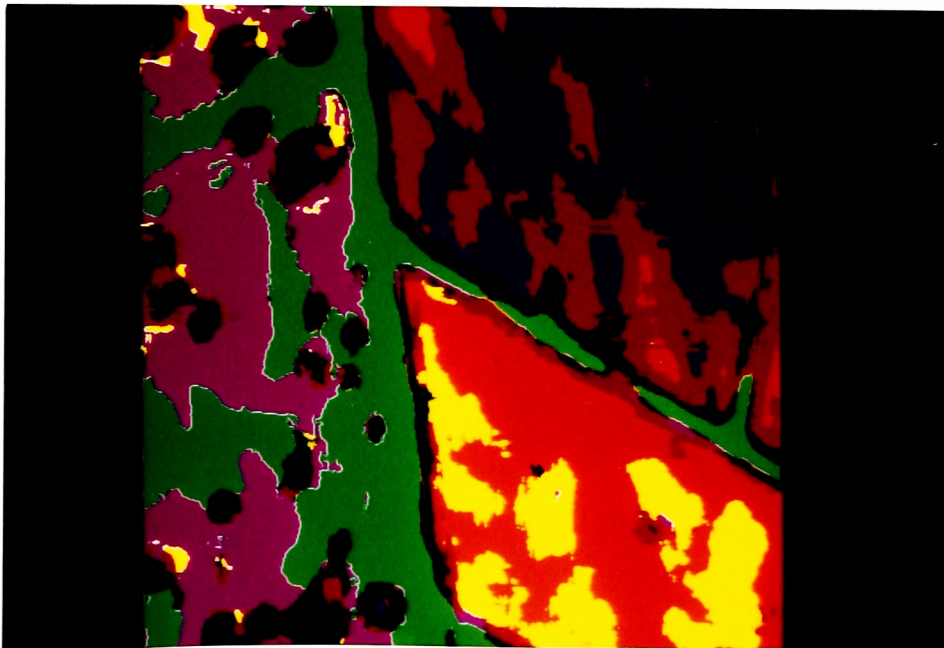
Figure   4.3-23

Gaussian Maximum Likelihood classification was carried out for analysis 21 after the training set was incremented by the textural feature image of feature entitled bright. Although this feature was not selected in final optimal feature space, computer classification showed the effects on global classification accuracy from it's addition to the analysis. Test data was selected and classified by the computer and identified by the analyst for determination of classification accuracy. Results are illustrated in Figure 4.3-24 and in Appendix B-5 for 8 textural feature image training set. The abbreviation ERDAS 7 FOFTI stands for computer classification of image data using the 7 textural feature images for classifier training (see textural feature images in previous paragraph). The abbreviation 7 FOFTI + Bright stands for computer classification of image data, using the same 7 textural feature images for classifier training but incremented by the textural feature image of feature entitled bright. The computer classified image is displayed in Figure 4.3-25.

ERDAS 7 FOFTI vs 7 FOFTI + Bright
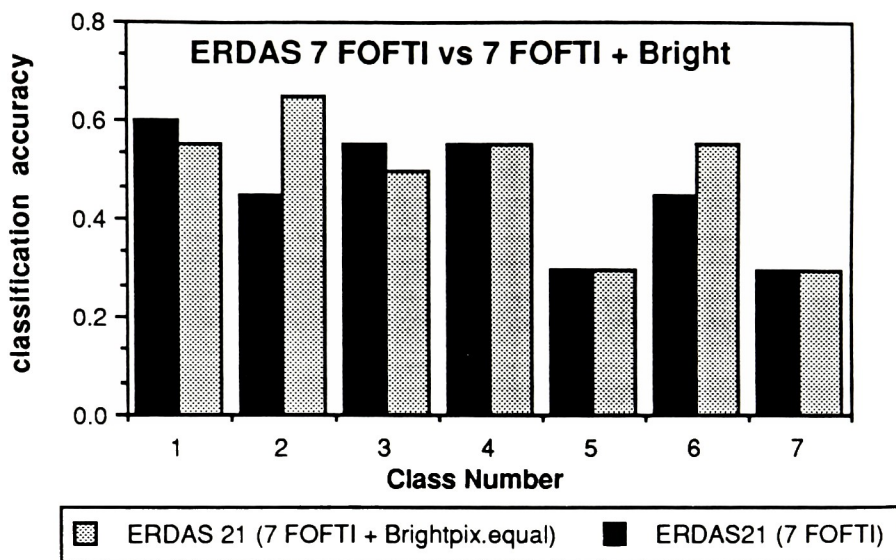
GCA Test Data ERDAS(7) : 45.7%   GCA Test Data ERDAS(8) : 48.6%

Figure   4.3-24
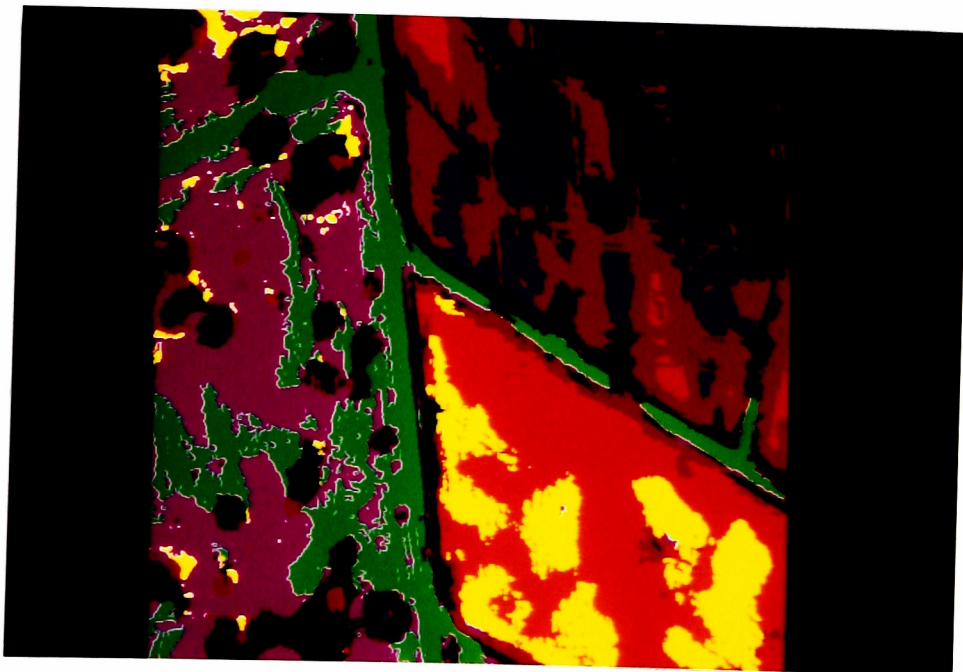
Figure   4.3-25

4.4   Processing of Image Forest

Five ground cover classes were identified within image **Forest** entitled : 1) light grey trees, 2) orchard, 3) light grey crops, 4) grey crops, and 5) dark grey crops.   All ground cover classes displayed overall uniform textural content over their digitized surfaces with exception to class entitled orchard where texture was irregular over large image areas.   Image Forest is shown in Figure 4.4-1.
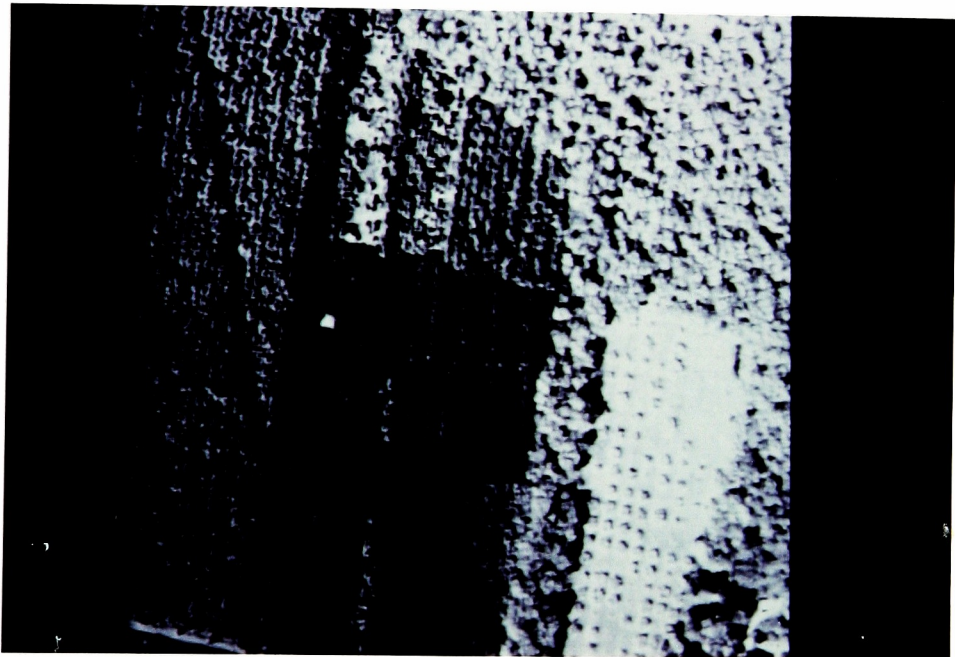
Figure   4.4.1

Similar processing was carried out on image Forest as was performed on both previous images. Five image class windows were selected across the five ground cover classes to encompass classifier training pixels.

4.4.1   Analyses 21, 25, 31, 3221, 3231, 31*r

Image Forest was processed in 6 ways entitled : 21, 25, 31, 3221, 3231, and 31*r as listed in Table 4.4-1.   Analyses (21, 25, 31) and (3221, 3231) illustrated the effects of increased image sampling window size and constant number of grey tone quantization levels on classification accuracy.   Analyses (21, 3221) and (31, 3231) demonstrated the effects of increased grey tone quantization level and constant image sampling window size.   Analysis 31*r displayed the effects of a 30 degree clockwise image rotation.

TABLE 4.4-1

Analysis Parameter Settings for Analyses 21, 25, 31, 3221, 3231, 31*r over Image Forest

	Analysis				
Parameters	**21**	**25**	**31/31*r**	**3221**	**3231**
Quantization Levels	16	16	16/16	32	32
Sampling Window Size	21*21	25*25	31*31	21*21	31*31
Interpixel Sampling	1	1	1	1	1
Degrees of Rotation	0	0	0/30 cw	0	0

note : cw = clockwise

Final optimal feature spaces were derived and tested for all analyses by classifying image training and test data. All processes are summarized in the paragraphs that follow.

The logic of Appendix E was followed for all analyses, with exception to analysis 31*r which was carried out in similar fashion to analysis 21*r on image Field. Forty-six textural features were measured for training pixels contained within each image class

window and are listed in Appendix A-1. The distribution of class training pixels per class for each analysis is listed in Table 4.4-2.

TABLE 4.4-2

Number of Training Pixels per Class for Analyses 21, 25, 31, 3221, 3231 for Image Forest

Analysis :	21/3221	25	31/3231
Image Class			
1) light grey trees	1075	819	495
2) orchard	1075	819	495
3) light grey crops	1075	819	495
4) grey crops	1075	819	495
5) dark grey crops	1075	819	495

For all analyses, class pooled covariance matrices of dimension 46*46 were derived, feature correlation coefficient matrices were extracted, and initial optimal features were derived from analysis of feature correlation coefficients with an established threshold value of 60%. Seventeen initial optimal features were derived for analysis 21, 17 features for analysis 25, 14 features for analysis 31, 16 features for analysis 3221, and 15 features for analysis

123

3231. For analysis 31, initial optimal features were measured for all pixels and individual textural feature images were derived of image Forest. Initial optimal feature spaces are listed in Appendix C-1 and textural feature images are displayed in Appendix B-7.

The training data files were restructured to contain the features of their initial optimal feature spaces and were processed to derive class mean vectors, individual class covariance matrices, optimal feature combination subsets and maximum divergence statistics. From analysis of maximum divergence statistics, final optimal feature spaces were established. Figures 4.4-2 through 4.4-6 show the dimensions of all final optimal feature spaces for analyses 21, 25, 31, 3221, and 3231 in terms of Maximum Divergence vs Number of Features in Final Optimal Feature Space. Circled data points represent these dimensions. Final optimal feature spaces are listed in Appendix C-1. Figure 4.4-7 illustrates the number of initial and final optimal features for all analyses. Best feature combinations for each analysis are listed in Appendix C-5.
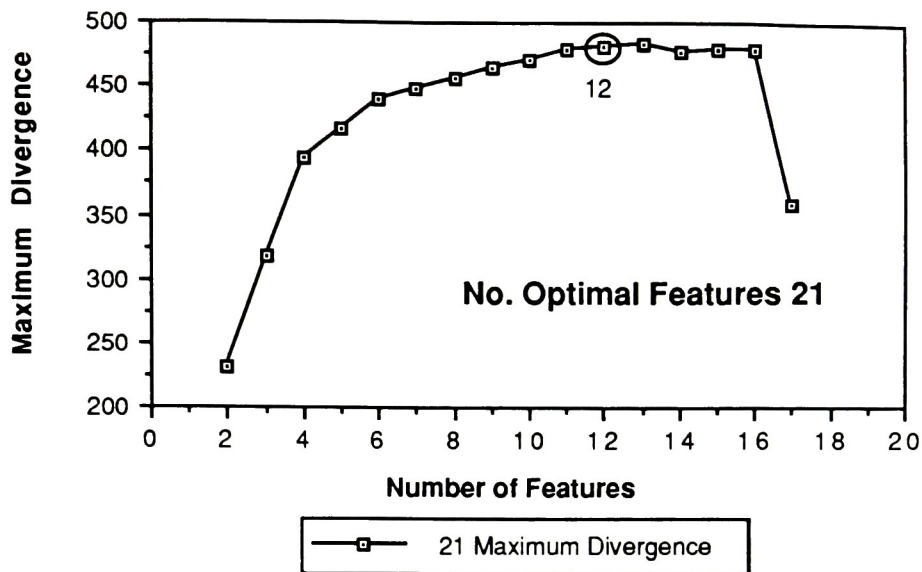
Figure   4.4-2
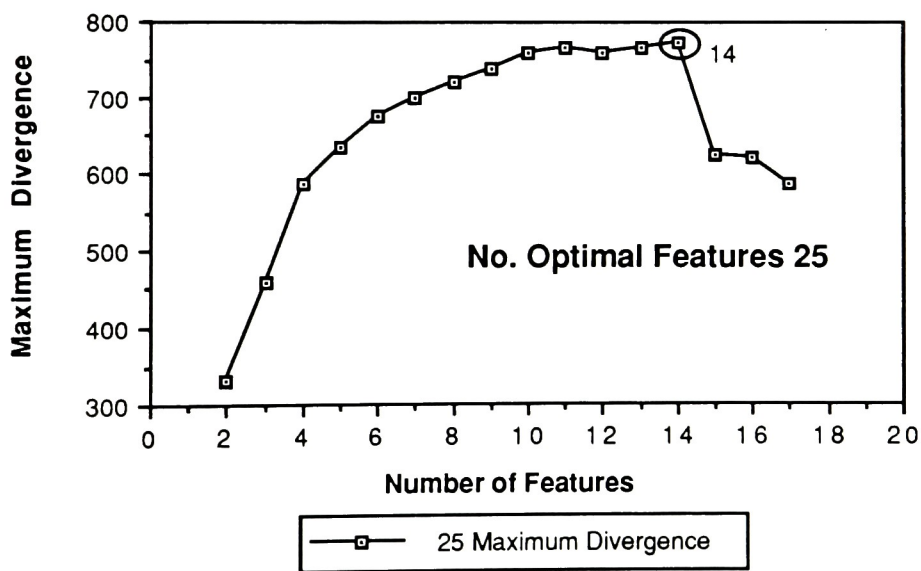
Figure   4.4-3
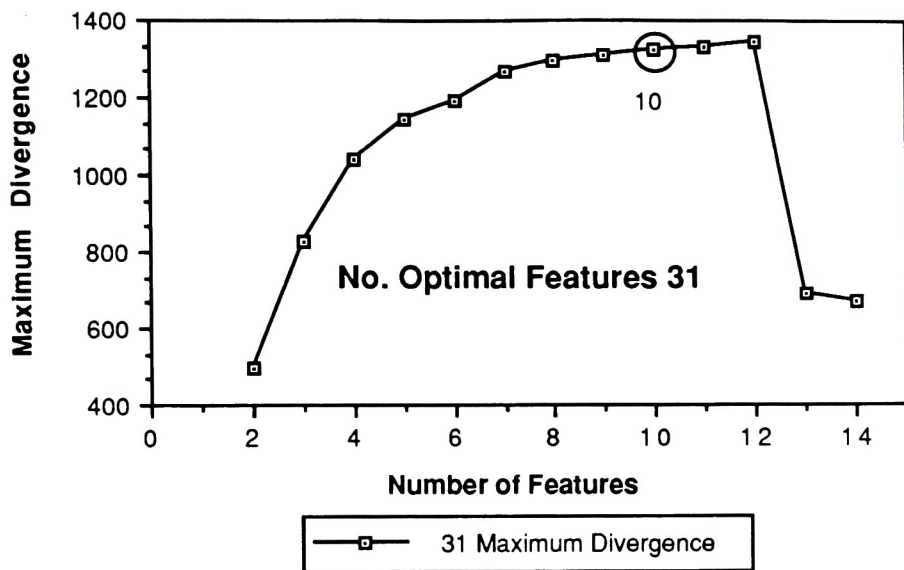
Figure   4.4-4
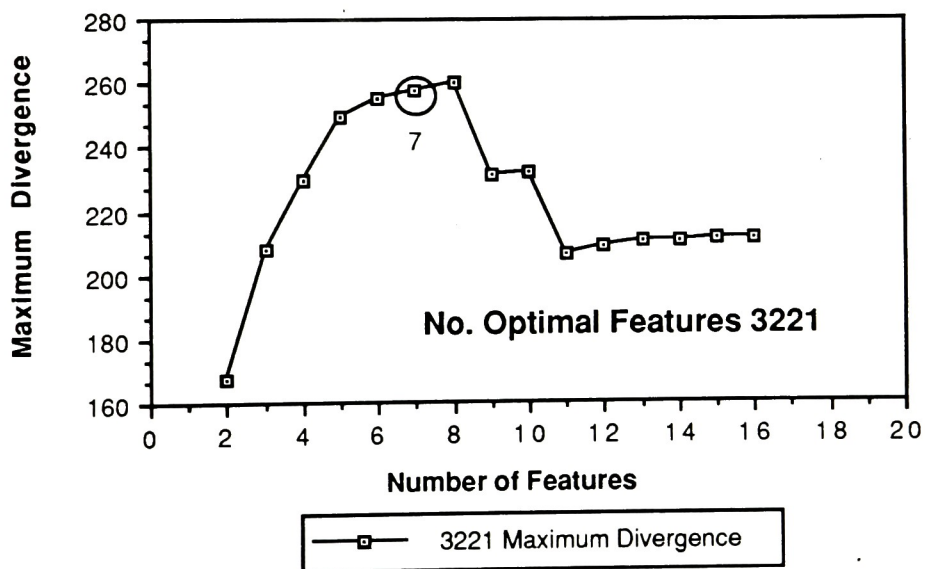
Figure   4.4-5
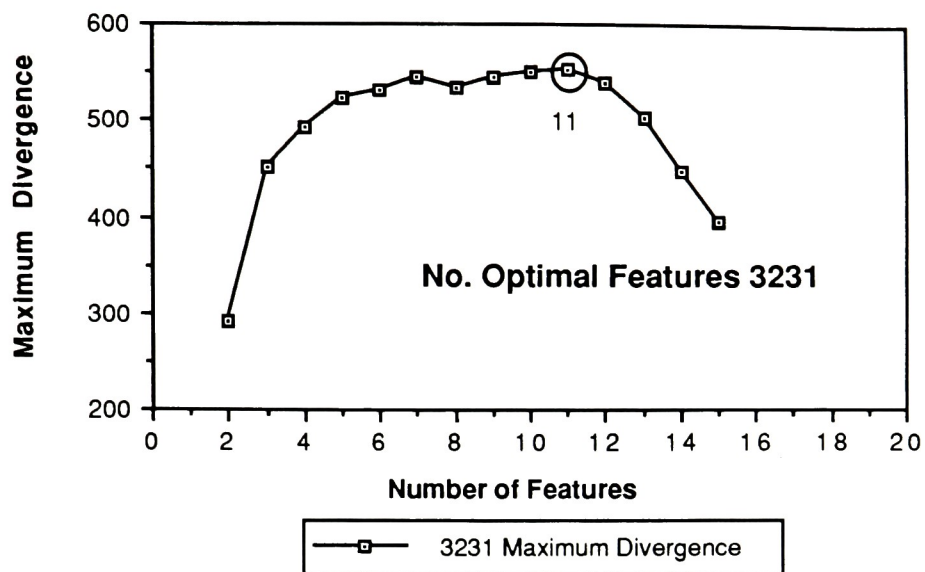
Figure   4.4-6

INITIAL AND FINAL OPTIMAL FEATURE SPACES


```
┌──────────┐      ┌──────────┐      ┌──────────┐
│    21    │      │    25    │      │    31    │
│  46 feat │      │  46 feat │      │  46 feat │
└──────────┘      └──────────┘      └──────────┘
        ┌──────────┐      ┌──────────┐
        │   3221   │      │   3231   │
        │  46 feat │      │  46 feat │
        └──────────┘      └──────────┘
    ↓        ↓        ↓        ↓        ↓
```

INITIAL OPTIMAL FEATURE SPACE

17f   16f   17f   15f   14features

FINAL OPTIMAL FEATURE SPACE

12f   7f   14f   11f   10features
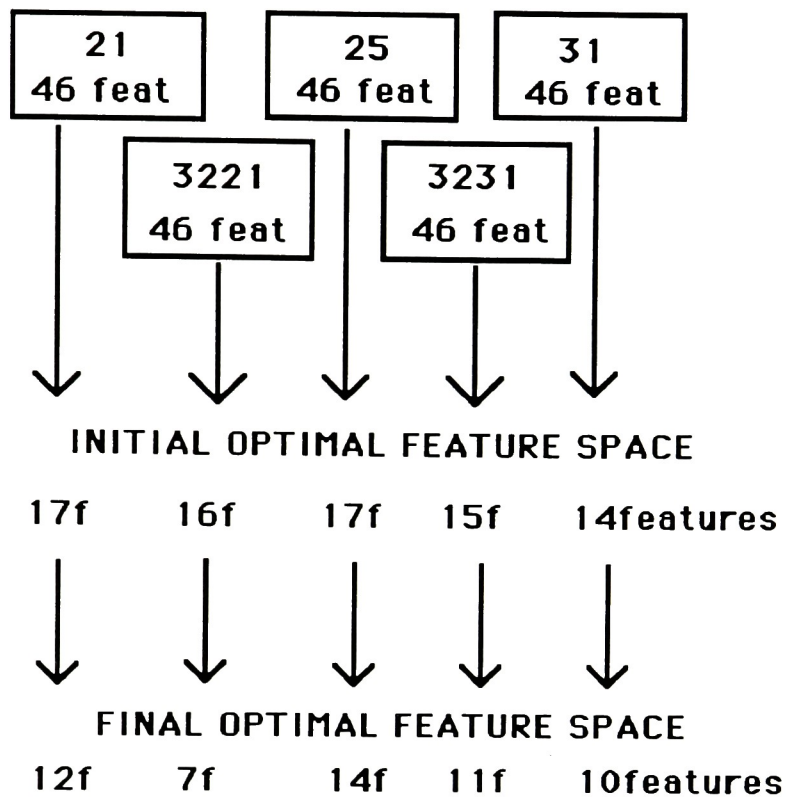
Figure 4.4-7

For each analysis, training data were reduced to include features of final optimal feature space.   Class statistics were derived and training and test data were classified.   Thirty test pixels per class were selected at random by the analyst for each analysis.   Results of classification are listed in Appendix C-2 and in Figures 4.4-8 through 4.4-12.

through 4.4-12. The term "Irfarm" in all plots relates to analyses over image Forest.

Plot Irfarm21
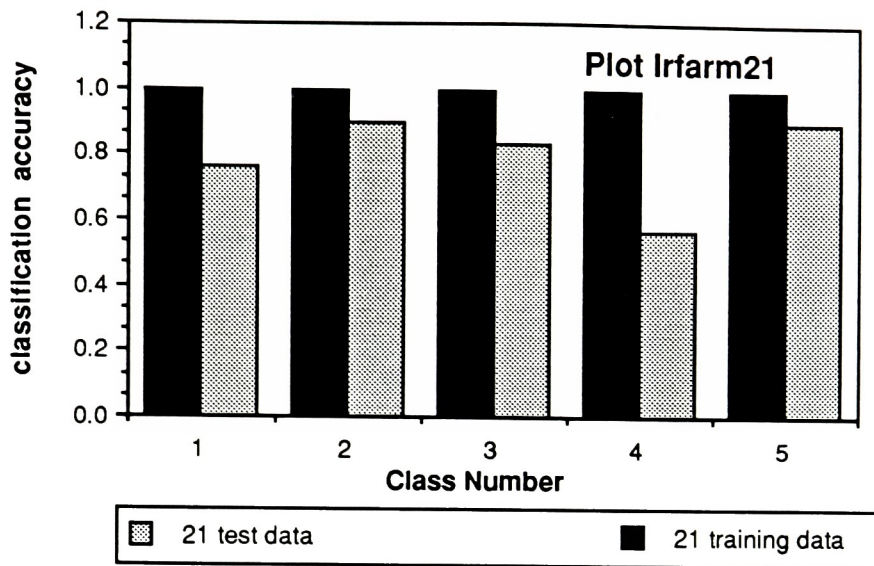
GCA 21 Training Data : 99.9%   GCA 21 Test Data : 79.3%

Figure   4.4-8

Plot Irfarm25
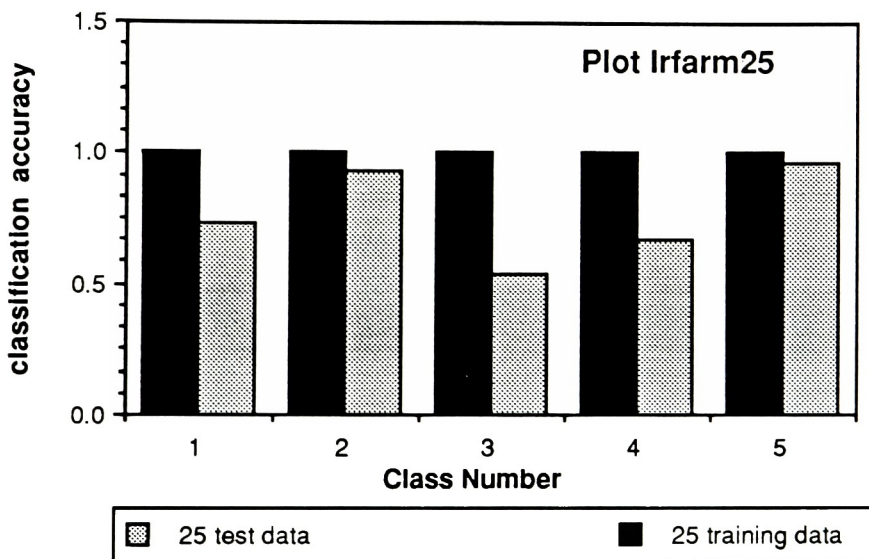
GCA 25 Training Data : 100.0%   GCA 25 Test Data : 76.6%

Figure   4.4-9

Plot Irfarm31
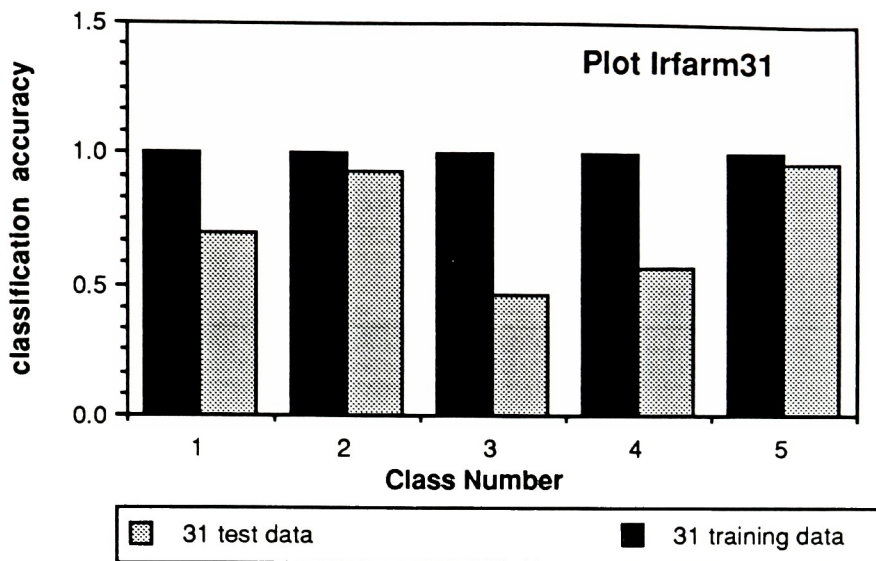
GCA 31 Training Data : 100.0%   GCA 31 Test Data : 72.6%

Figure   4.4-10

Plot Irfarm3221
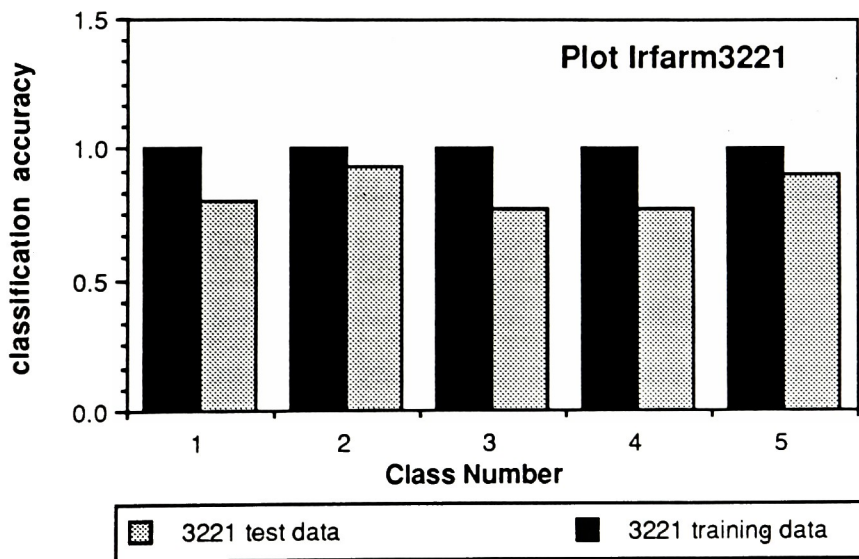
GCA 3221 Training Data : 100%   GCA 3221 Test Data : 83.3%

Figure   4.4-11

Plot Irfarm3231

Legend: 3231 test data | 3231 training data
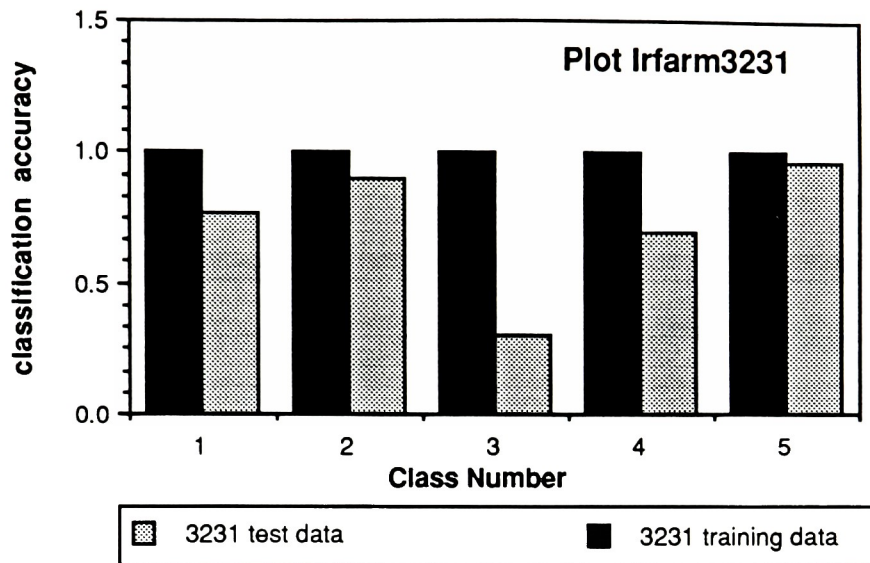
GCA 3231 Training Data : 100%   GCA 3231 Test Data : 72.6%

Figure   4.4-12

4.4.2   Individual Class Covariance Matrices vs Pooled Class Covariance Matrix

For all analyses, final optimal feature selection and classification exercises were performed using individual class covariance matrices over class pooled covariance matrices. This ensured optimal process accuracy. Statistical hypothesis testing of the inequality of individual class covariance matrices also excluded pooled class covariance matrices from consideration. Section 4.2.3 provides further argument.

4.4.3   Analysis 31*r

For analysis 31*r, 30 test pixels were selected at random by the analyst within each of the five image classes of the rotated image. Each test pixel was characterized by final optimal features of analysis 31, and was derived from image sampling windows of dimension 31*31, oriented at 0 degrees orientation with respect to the image. Classification results of test data from analyses 31 and 31*r are listed in Appendix C-2 and are illustrated in Figure 4.4-13.
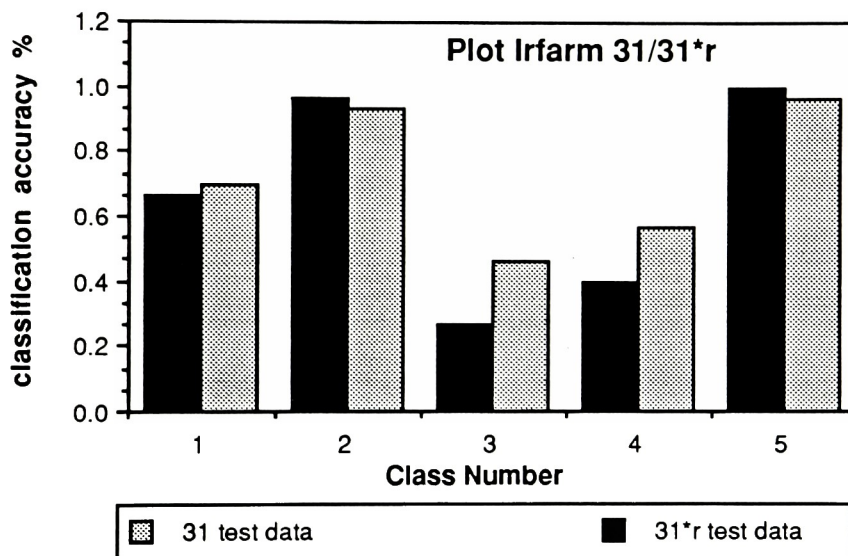
GCA 31 Test Data :   72.6%   GCA 31*r Test Data :   65.9%

Figure   4.4-13

4.4.4 Validation of Final Optimal Feature Space Construction Process

The optimality of final optimal feature spaces was tested for analyses 21, 25, and 31. Subsets #1 and #2 were established for each analysis, containing 8 features apiece for analysis 21, 8 features apiece for analysis 25, and 7 features apiece for analysis 31. Subsets #1 contained features which were deemed optimal from their respective initial optimal feature spaces by discriminant analysis while subsets #2 contained an equal number of remaining features from the initial optimal feature spaces. Individual feature subsets for analyses 21, 25, and 31 are listed in Appendix C-3.

Training and test data files were restructured for each analysis to include the features of their respective subsets. Individual class statistics required to train Gaussian Maximum Likelihood classifiers were derived, and image training and test data were classified. Results from these classification exercises are listed in Figures 4.4-14 through 4.4-19 and in Appendix C-4. Figure 4.2-24 illustrates this process.

Plot Irfarm21 Subsets (training data)
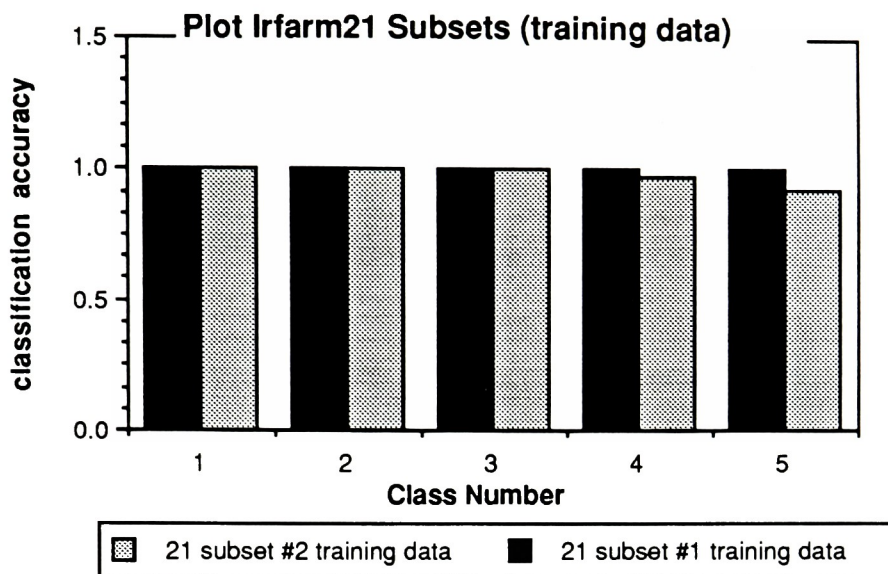
GCA 21 Subset #1 : 100.0%   GCA 21 Subset #2 : 98.0%

Figure   4.4-14

Plot Irfarm21 Subsets (test data)
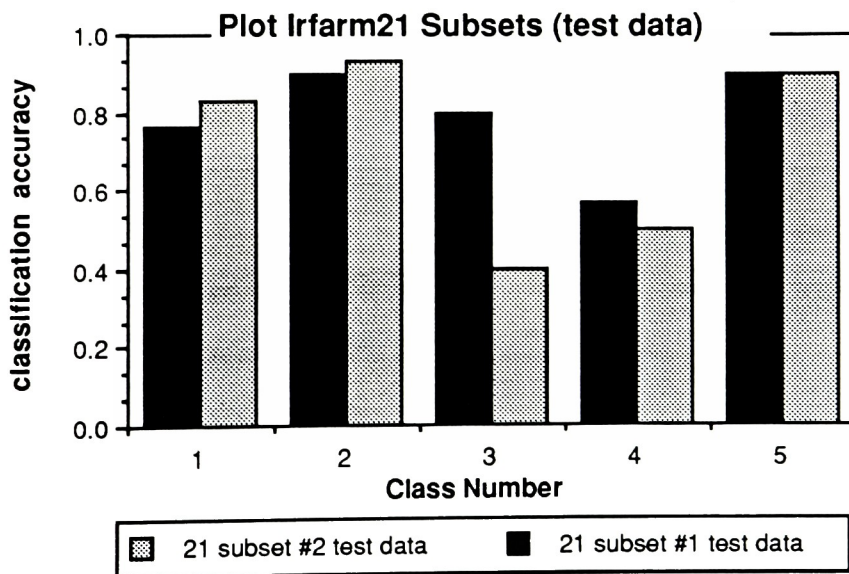
GCA 21 Subset #1 : 78.6%   GCA 21 Subset #2 : 71.3%

Figure   4.4-15

Plot Irfarm25 Subsets (training data)
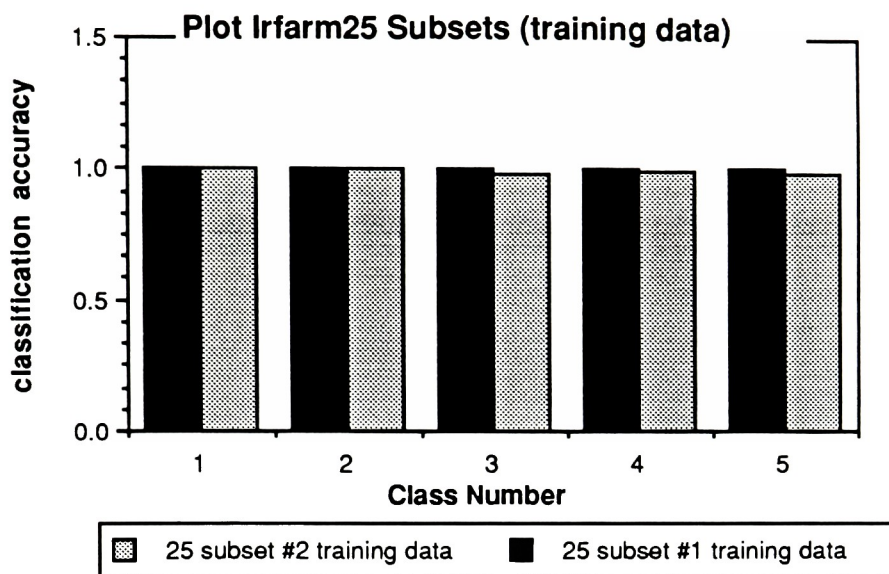
GCA 25 Subset #1 : 100.0%   GCA 25 Subset #2 : 99.3%

Figure   4.4-16

Plot Irfarm25 Subsets (test data)
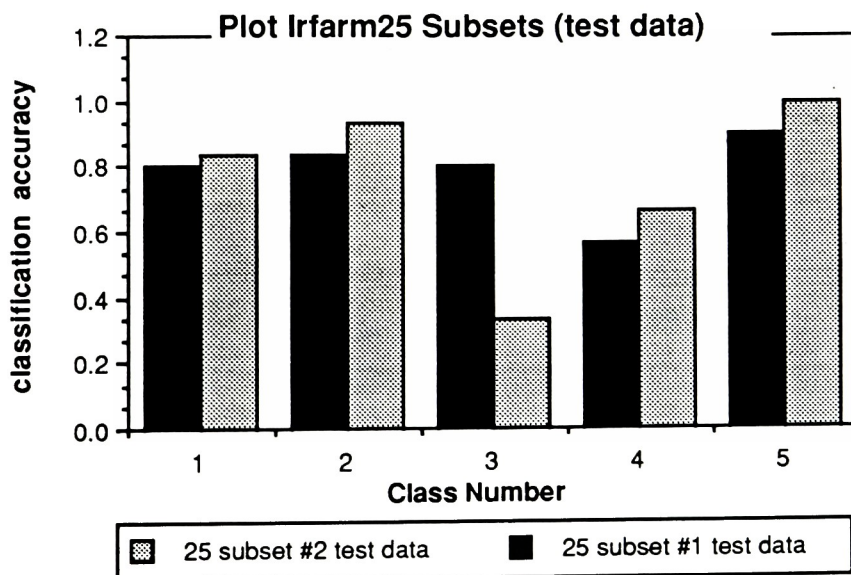
GCA 25 Subset #1 : 78.0%   GCA 25 Subset #2 : 75.2%

Figure   4.4-17

Plot Irfarm31 Subsets (training data)
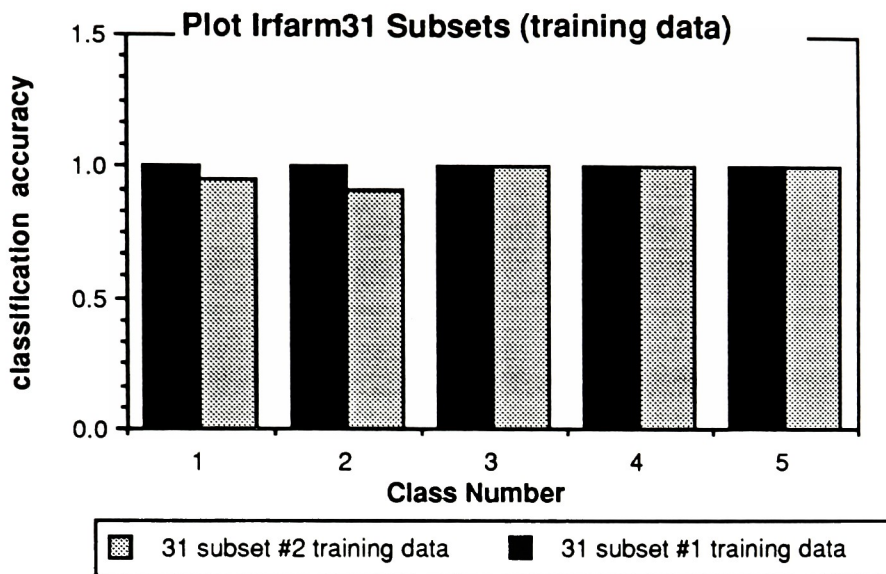
GCA 31 Subset #1 : 100.0%   GCA 31 Subset #2 : 97.3%

Figure   4.4-18

Plot Irfarm31 Subsets (test data)
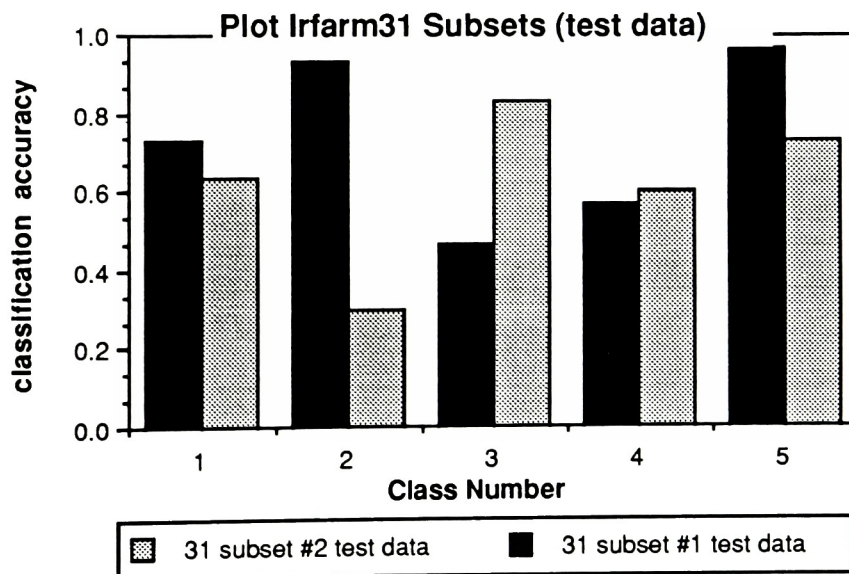
GCA 31 Subset #1 : 73.3%   GCA 31 Subset #2 : 62.0%

Figure   4.4-19

4.5 Robustness Testing of Classification Accuracies

To ensure that classification accuracies were relatively constant over all analyses for all images, analysis 31 from image Forest was subjected to 4 additional classification tests.   Each test was performed on dissimilar sets of 30 test pixels, selected at random by the analyst, and classified in final optimal feature space from analysis 31.   Global classification accuracies of test data sets were 76.00%, 70.66%, 73.33%, and 71.33% compared to 72.66% accuracy from the original set of test data.   Despite this relative consistency in results, it is premature to establish trends (i.e. increases, decreases, constant readings) in classification accuracies from changes in feature parameter settings for all analyses.   This is because classification accuracies are also dependant on the textures which are measured and are changing from analysis to analysis. Observed behaviour in classification accuracies for all analyses are listed in Appendix F.

4.6 Processing Times for all Analyses

The computer processing times (CPU times) required for execution of all analyses were of considerable length.   Derivation of forty-six textural features for all training pixels required CPU times ranging from 3 to 7 hours, depending on the number of training pixels selected from all classes.   Derivation of run-length type textural

features required significantly more CPU time than that required of other features.

For multivariate discriminant analysis of initial optimal feature spaces, CPU times ranged between 4 hours to several days. More commonly encountered CPU times for this process were approximately 10 hours for derivation of best final optimal features from initial optimal feature spaces containing between 10 and 15 features. All remaining feature selection and classification exercises required CPU times ranging between a few minutes to approximately half an hour.

The processing times required to generate textural feature images ranged from 18 hours to 10 days. Textural feature images of run-length features were those which required excessively long processing times. This explained the long hours required to generate initial textural feature data files for training data.

5.0 Discussion of Results

The following section consists of a discussion of the method of analysis, of the results from all analyses.

5.1 Discussion of the Method of Analysis

<u>Image Preprocessing</u> : All images were subjected to histogram equalization and histogram level slicing to ensure optimal image quality and maximum accuracy of feature selection and classification exercises. Histogram equalization removed unwanted contrast and brightness levels from the image. Histogram level slicing eliminated unwanted image data and increased distinction between image textures. Absence of other apparent radiometric or geometric inconsistencies excluded use of other preprocessing algorithms.

All analyses were performed on monochrome images equal-probability quantized to either 16 or 32 brightness levels. Quantization levels were chosen based on classification experiments carried out by other researchers where similar parameter settings were used.[44] Selection of sufficient quantization levels ensured minimal loss of class textures, reliable classification of image training and test data and accurate selection of optimal features.

Training and Test Data : For analyses on images Urban and Field, several classes were allocated insufficient number of training pixels (for image Urban : analyses (162, 322) and (164, 324) for class entitled medium gray water, and analyses (164, 324) for class entitled medium grey water; for image Field : analyses (25, 25*5, 3225) for class entitled dark gray desert). This was caused by the small sizes of classes and to the large size of image class and sampling windows. Unreliable class statistics were derived and poor classification accuracies resulted.

Test data selected by the analyst included no mixed pixels to maximize classification accuracy. To render the classification process more representative, test data was selected at random by the computer and classified in optimal feature space.

Dimensions of Image Class and Sampling Windows : Image class windows were rectangular or square in shape and oriented at O degrees rotation with respect to the image. Image sampling windows were square in shape and of constant size for each analysis. Because of these restrictions it was difficult to position and size all windows to surround representative class textures and to capture sufficient numbers of training pixels. Image sampling window size was to capture the coarsest class texture, while ensuring accurate measurement of finer class textures.

<u>Derivation of Inverses for Individual Class Covariance Matrices</u> : The Moore-Penrose generalized inverse was calculated for singular individual class covariance matrices which existed when features displayed linearly interdependency.   Increasing the number of training pixels per class reduced linear interdependency by providing accurate class statistics and individual class covariance matrices more susceptible to being non-singular in nature.

The Moore-Penrose generalized inverse is derived from a slightly modified, non-singular version of the original matrix.   The slightly inaccurate inverse decreases classification and feature selection accuracy by virtue of its nature.   Moore-Penrose generalized inverses of individual class covariance matrices were recognized by the uneven shape of plots of Maximum Divergence vs Number of Features in Final Optimal Feature Space   Optimal feature selection and classification exercises were characterized by the even shape of these plots.

An alternative to use of Moore-Penrose generalized inverses would be to disregard feature spaces characterized by singular individual class covariance matrices as singularity reflects high correlation among features.   To minimize correlation, a higher threshold to derive initial optimal features from correlation coefficients would be established.   The Moore-Penrose approach was adopted as individual class covariance matrices frequently displayed

singularity and unacceptable large numbers of feature spaces would have been disregarded.

Derivation of Initial Optimal Feature Spaces by Analysis of Correlation Coefficient Matrices : The purpose of this exercise was to provide for significant increases in the efficiency of the feature selection process and for considerable reduction in the magnitude of initial optimal feature sets. Features correlated among themselves, beyond an established threshold, were removed from analysis as they contained redundant information.

By decreasing the threshold, a greater number of initial features would be discarded, discriminant analysis feature selection efficiency would be increased, and global classification accuracy would decrease. Increasing the value of this statistic would produce exact opposite results. The value of the threshold is function of classifier accuracy versus feature selection efficiency. Prior knowledge of the ability of features to provide for superior classification accuracy would allow for rejection of features which contribute little to process accuracy.

Selection of a feature from a correlated set, based upon maximum value of a derived correlation coefficient (see Section 2.5.1), did not guarantee its optimality. Analyses in Section 4.2.2 over image Urban classified image data within feature spaces which excluded final

optimal features and comparable classification accuracies were obtained. The feature spaces were derived by analysis of correlation coefficients and discriminant analysis. Since discriminant analysis ensured final optimal feature selection, analysis of correlation coefficients was proven a sub-optimal technique for selection of initial optimal features. This method of feature selection decreased the accuracy of the classification process but significant increased the efficiency of feature selection exercises.

Optimal Method for Formulation of Final Optimal Feature Spaces :
Optimal feature selection consists of performing multivariate discriminant analysis over all features. This approach was not carried out as excessive computation time would be required due to the high number of features to be considered. However, optimal feature selection was performed from subsets of features which were derived by multivariate discriminant analysis from their initial optimal feature spaces (see Sections 4.2.4, 4.3.4, and 4.4.4). Classification accuracies for these subsets were consistently superior to those for subsets containing other features from the initial optimal feature spaces

.

Derivation of Final Optimal Feature Spaces by Analysis of Maximum Divergence Statistics : Final optimal feature spaces were derived from multivariate discriminant analysis of initial optimal feature spaces. By decreasing the value of the threshold of 1% between

increments in divergence for increments in the dimensionality of final optimal feature space, classification accuracy would increase feature selection efficiency would decreases. Increases in the threshold would produce opposite effects.

All weights w_{ij} assigned to maximum divergence Mahalanobis-like distances d_{ij}, were equal to 1 for all analyses. It was of concern however, should a few Mahalanobis-like distances prove to be excessively large, that final optimal feature spaces would possess maximum value of maximum divergence statistic without ensuring maximum discrimination between all classes. A solution to this problem would consist of assigning appropriate weights to individual Mahalanobis-like distances based on some a priori knowledge of the distances between classes. Assignment of separate weights would also be performed when maximum classification accuracy of data into select classes is desired. Small weights would be assigned to Mahalanobis-like distances between classes where classification accuracy is unimportant, while larger weights would be assigned in an opposite manner.

Parameter Selection for Derivation of Textural Features : Optimal classification accuracy was possible when textural features provided for effective measurement of the textural information of underlying classes and provided for discrimination between image pixels. Code restrictions constrained parameter settings to be equal

over all classes for each analysis. Therefore, it was common practice to choose their values to capture the coarsest class texture. This would also allow for capture of finer textured classes.

Utilization of Pooled Class Covariance Matrices vs Individual Class Covariance Matrices : Statistical tests disallowed use of pooled class covariance matrices in final optimal feature selection and classification exercises. Maximum process accuracy was thereby ensured. Pooled class covariance matrices were used for derivation of initial optimal feature spaces to achieve significant increases in the efficiency of the feature selection process. As stated in Section 2.5.1 a more accurate version of a pooled covariance matrix could be derived from all pixel observations by considering them all from one class. This approach is more precise and warrants attention in additional studies.

5.2 Analysis of Results

Results of all analyses for all images are discussed in the following section. All results show the effects of feature selection, image class and sampling window size, and feature parameter settings on classification accuracy

5.2.1 Analysis of Results for Image Urban

Correct dimensioning and positioning of image class and image sampling windows was required to ensure accurate classifier training. The small size of classes light grey water (1), medium grey water (2), dark grey water (3), and light grey grass (8) restricted the dimensions of their image class windows, the number of training pixels per class selected therein, and the size of all image sampling windows. Irregular shapes of all classes of water (1, 2, 3), and non-uniform textures across classes residential (4) and dark grey forest (6) affected positioning of image class windows and restricted capture of class textures by image sampling windows. Three by three and five by five pixel image sampling windows were selected to capture training pixels from all classes for all analyses. Of concern was the ability of image sampling windows to capture the texture of coarser textured classes such as residential (4) and orchard (7).

147

Global classification accuracies for all analyses ranged between 83.4% - 91.1% for training data, and between 49.4% - 63.9% for test data. Results are listed in Table 5.2.1-1 and in Section 4.2.1. All analysis designates are characterized by sampling window size/quantization level attributes between parentheses.

TABLE 5.2.1-1

Global Classification Accuracy for Analyses 162, 164, 322, 324 over Image Urban

	Global Classification Accuracy	
Analysis	Training Data	Test Data
162 (3x3/16q)	83.4%	59.0%
164 (5x5/16q)	91.1%	51.5%
322 (3x3/32q)	85.1%	63.9%
324 (5x5/32q)	85.0%	49.4%

Small global classification accuracy of test data compared to training data was due to poor classification of test data in all classes of water (1, 2, 3), light grey grass (8), dark grey forest (6), and highway (11). Appendix A-3 lists all classification accuracies for all classes. Low classification accuracies were result of inaccurate derivation of class statistics from insufficient and unrepresentative training pixels. This was due to improper positioning of image class windows, to improper size of image sampling windows, and to the inability of feature parameter settings to ensure discriminant measures of class textures.

Classification of image test data revealed that increasing image sampling window size from 3*3 to 5*5 pixels, for constant quantization level (16 or 32 grey tones) decreased global classification accuracy from 59.0% to 51.5% for analyses 162 and 164, and from 63.9% to 49.4% for analyses 322 and 324. Increases in quantization levels from 16 to 32, at constant image sampling window size, resulted in marginal changes in classification accuracy from 59.0% to 63.9% for analyses 162 and 322, and from 51.5% to 49.4% for analyses 164 and 324.

For different feature parameter settings, classification accuracies are function of class textures. Classification accuracy is function of the ability of the analyst to select correct feature parameters for all textures to ensure maximum discrimination between class

pixels. For class dark grey forest (6) over all analyses, inaccurate feature parameter settings contributed to decrease classification accuracy. This argument applies for all classes were poor results were obtained.

For analyses (162, 261), (164, 461), (322, 223), and (324, 423), classification accuracy in final optimal feature space (optimal) was measured against classification of in sub-optimal feature space (sub-optimal) which excluded previous final optimal features. Section 4.2.2 provides further explanation. Classification accuracies demonstrated the next to optimal characteristic of the feature selection technique for derivation of initial optimal features. Table 5.2.1-2 summarizes results from all analyses.

TABLE 5.2.1-2

Global Classification Accuracy for Analyses (162, 261), (164, 461), (322, 223), (324, 423) over Image Urban

---

		Global Classification Accuracy	
	Analysis	Training Data	Test Data
Optimal	162 (3x3/16q)	83.4%	59.0%
Sub-Optimal	261 (3x3/16q)	78.3%	61.4%
Optimal	164 (5x5/16q)	91.1%	51.5%
Sub-Optimal	461 (5x5/16q)	83.2%	45.1%
Optimal	322 (3x3/32q)	85.1%	63.9%
Sub-Optimal	223 (3x3/32q)	85.5%	63.9%
Optimal	324 (5x5/32q)	85.0%	49.4%
Sub-Optimal	423 (5x5/32q)	95.8%	58.2%

- -

Global classification accuracies demonstrated that sub-optimal final feature spaces for analyses 261, 461, 223, and 423 performed as well as optimal final optimal feature spaces for analyses 162, 164, 322, and 324. As discriminant analysis, used to derive final optimal feature spaces, was an optimal process, initial optimal feature selection by analysis of correlation coefficient matrices was determined a sub-optimal process.

For analyses 162, 164, 322, and 324, individual and pooled covariance matrices were compared for final optimal feature selection and classification of image data. Results are listed in Table 5.2.1-3 and in Section 4.2.3.

TABLE 5.2.1-3

Global Classification Accuracy for Analyses
162, 164, 322, 324 using Pooled and Individual Class
Covariance Matrices for Training Data over Image Urban

| Analysis | Global Classification Accuracy | |
	Individual Cov. Matrices	Pooled Cov. Matrix
162 (3x3/16q)	83.4%	54.6%
164 (5x5/16q)	91.1%	69.5%
322 (3x3/32q)	85.1%	60.8%
324 (5x5/32q)	85.0%	52.0%

Global classification accuracies using pooled class covariance matrices were consistently inferior to those obtained using individual class covariance matrices. Pooled class covariance matrices increased the efficiency of the feature selection process but decreased the accuracy of classification exercises. As process accuracy was vitally important pooled class covariance matrices were not used.

The process by which final optimal feature spaces were derived was validated for analyses 162, 164, and 322 by performing classification over subsets of features   Section 4.2.4 provides further explanation.   Two subsets of features were derived from initial optimal feature space; subset #1 by discriminant analysis and subset #2 by random selection of remaining features.   Results are listed in Table 5.2.1-4.

TABLE 5.2.1-4

Global Classification Accuracy for Analysis 162, 164, 322, 324 Subsets' #1 and #2   over Image Urban

	Global Classification Accuracy			
	Training Data		Test Data	
Analysis	Subset #1	Subset #2	Subset #1	Subset#2
162 (3x3/16q)	82.6%	37.0%	62.4%	29.4%
164 (5x5/16q)	91.5%	78.5%	49.3%	45.4%
322 (3x3/32q)	83.6%	44.4%	64.5%	37.8%

Global classification accuracies for subsets #1 were proven consistently superior to those of subsets #2. The classification results validated the process of final feature selection from initial optimal feature spaces.

Initial and final optimal features for analyses 162, 164, 322, and 324 were not of singular occurrence. Initial optimal features angsecaver, angsecrang, contraver, corraver, varraver, varrang, sumenrang, diffenrang, and meascoranga were common to all analyses. Final optimal features angsecaver, varraver, and diffrang were also common to all analyses while final optimal features angsecrang, contraver, corraver, indiffrang, sumenrang, entrang, and meascoranga belonged in two or three analyses. Appendix A-2 lists all final and initial optimal features for analyses over image Urban.

5.2.2   Analysis of Results for Image Field

Uniform class textures across large areas for image Field allowed positioning and dimensioning of image class and sampling windows to capture sufficient training pixels. Large image sampling windows allowed for precise recording of class textural information and for accurate estimation of class statistics. Global classification accuracies for all analyses are listed in Table 5.2.2-1 and in Section 4.3.1

TABLE 5.2.2-1

Global Classification Accuracy for Analyses
17, 21, 25, 21*3, 25*5, 3225 over Image Field

Analysis	Global Classification Accuracy	
	Training Data	Test Data
17 (17x17/16q)	98.7%	67.6%
21 (21x21/16q)	99.4%	66.2%
25 (25x25/16q)	99.5%	63.3%
21*3 (21x21/16q)	99.9%	64.7%
25*5 (25x25/16q)	99.9%	64.3%
3225 (25x25/32q)	100.0%	55.2%

For analyses 17, 21, 25, 21*5, and 3225, global classification accuracies for training data ranged between 98.7% and 100.0%, and for test data between 63.3% and 67.6%. The exception was for analysis 3225 where global classification accuracy for test data

was 55.2%. The decrease in global classification accuracy between training and test data was due to insufficient numbers and non-representativeness of training data for several classes. Inadequate size of image class and image sampling windows, the small dimension and awkward shape of classes light grey grass (5) and dark grey desert (6), and inappropriate feature parameter settings all contributed to decreasing classification accuracy. Reductions in classification accuracy from test data compared training data was apparent for classes light grey desert (1), light grey grass (5), and dark grey desert (6) for several analyses. Appendix B-2 lists classification accuracies for all classes.

Global classification accuracies of image training and test data were compared for analyses (17, 21, 25) to demonstrate the effects of constant number of quantization levels and increased image sampling window size. Global classification accuracy decreased for test data from 67.6% to 63.3% for increased image sampling window size from 17*17 to 25*25 square pixels. Global classification accuracies were compared for analyses (25, 3225) to demonstrate the effects of constant image sampling window size and increasing quantization levels from 16 to 32. Global classification accuracy of test data decreased from 63.3% to 55.2% for analyses 25 and 3225 respectively. These results confirm that classification accuracies are function of the textural content of each class under analysis for constant feature parameter setting.

The effect of a 30 degrees clockwise rotation on global classification accuracy was illustrated in analysis 21*r. Results from analyses 21 and 21*r are listed in Table 5.2.2-2 and in Section 4.3.3.

TABLE 5.2.2-2

Global Classification Accuracy for Analyses 21, 21*r over Image Field

Analysis	Global Classification Accuracy Test Data
21 (21x21/16q)	66.2%
21*r (21x21/16q)	64.7%
(30 degree clockwise rotation)	

Global classification accuracies for analyses 21*r and 21 were similar. This was due to the shift-invariant property of all textural features, as was reported in Section 2.3.

The effect of image magnification on global classification accuracy was illustrated in analyses 21*3 and 25*5. Global classification accuracies for analyses (21, 21*3) were comparable, as were those for analyses (25, 25*5). The small scale of image Urban contributed to the insensitivity of classification accuracy from increased magnification. Larger increases in image magnification would probably produce different classification accuracies. Table 5.2.2-1 lists results.

The process where by final optimal feature spaces were derived was tested for analyses 21 and 25. This was carried out by performing classification exercises of similar training and test data over subsets of features of equal number derived from their initial optimal feature spaces. Subset #1 features were derived by discriminant analysis and subset #2 features were chosen at random from those remaining. Results are listed in Table 5.2.2-3 and in Section 4.3.4.

TABLE 5.2.2-3

Global Classification Accuracy for Analyses 21, 25 Subsets' #1 and #2 over Image Field

---

	Global Classification Accuracy			
	Training Data		Test Data	
Analysis	Subset #1	Subset #2	Subset #1	Subset#2
21 (21x21/16q)	99.7%	97.2%	64.3%	58.5%
25 (25x25/16q)	99.1%	87.8%	62.8%	46.6%

Global classification accuracies for subsets #1 were proven consistently superior to those for subsets #2. These results confirmed the optimality of discriminant analysis for derivation of final optimal features from initial optimal spaces for all analyses.

Computer selection of test pixels decreased global classification accuracies from 67.6% to 49.3% for analysis 17 and from 66.2% to 55.0% for analysis 21 (see Section 4.3.2). These results were expected since computer selected test pixels included mixed pixels

160

which were often misclassified due to their ambiguous nature. Test pixels selected by the analyst contained no mixed pixels.

Computer executed Gaussian Maximum Likelihood classification of image Field for analysis 21 was carried out using textural feature images of features angsecaver, angsecrang, corraver, corrang, varrang, moyenne, and ecart to train the classifier. Global classification accuracy decreased to 45.7% from 55.0% (obtained in final optimal feature space). Such differences are attributable to the number and location of different training pixels used to train each analysis classifier.

Identical classification was carried out for analysis 21 for training set incremented by the textural feature image derived from feature entitled bright. Global classification accuracy increased from 45.7% to 48.6%. Addition of textural feature image Brightpix.equal to final optimal feature space provided for an increase in the distinction between image classes as is seen in Figure 4.3-24.

Initial optimal features brightness, average, variance, angsecaver*, angsecrang*, corraver*, corrang, varrang*, indiffrang, sumenrang*, diffenrang*, meascoranga*, meascoraverb, sreimrang and gradnt were common to 4 or more of the six analyses 17, 21, 25, 21*3, 25*5, and 3225. These features were also final optimal features for 4 or more of these analyses, with exception to features brightness

and gradnt.   All initial optimal features annotated by an asterisk *
were also initial optimal features for analyses 162, 164, 322, and
324 over image Urban.   Appendix B-1 lists all final and initial
optimal features for analyses of image Field.

5.2.3   Analysis of Results for Image Forest

Large image classes, and uniform and distinct class textures
facilitated accurate dimensioning and positioning of image class and
sampling windows.   Accurate measurement of class textures from
within image sampling windows was possible to ensure maximum
discrimination between classes in feature space.   Large numbers of
training pixels were extracted which allowed for precise estimation
of class statistics. Global classification accuracy for all analyses
are listed in Table 5.2.3-1

162

TABLE 5.2.3-1

Global Classification Accuracy for Analyses
21, 25, 31, 3221, 3231 over Image Forest

Analysis	Global Classification Accuracy	
	Training Data	Test Data
21 (21x21/16q)	99.9%	79.3%
25 (25x25/16q)	100.0%	76.6%
31 (31x31/16q)	100.0%	72.6%
3221 (21x21/32q)	100.0%	83.3%
3231 (31x31/32q)	100.0%	72.6%

For analyses 21, 25, 31, 3221, and 3231, global classification accuracies for training data ranged between 99.9% and 100.0% and for test data between 72.6% and 83.3%. The decrease in global classification accuracy between training and test data was attributed to inaccurate estimation of class statistics due to lack of representative training data for individual classes, and to the

inability of feature parameter settings to provide for optimal discriminant measures of class textures.

Global classification accuracies were compared for analyses (21, 25, 31) and (3221, 3231) to show the effects from constant number of quantization levels and increased image sampling window size. Global classification accuracies of test data for analyses (21, 25, 31) for 16 grey tone quantization levels, decreased from 79.3% to 72.6% for increased image sampling window size from 21*21 to 31*31 square pixels. For analyses (3221, 3231) similar effects were observed for a constant 32 quantization levels. Global classification accuracies for test data decreased from 83.3% to 72.6% for increased image sampling window size from 21*21 to 31*31 pixels. Global classification accuracies were compared for analyses (21, 3221) and (25, 3225) to demonstrate the effects of constant image sampling window size and varying number of quantization levels. Global classification accuracy of test data increased from 79.3% for analysis 21 to 83.3% for analyses 3221 while a constant classification accuracy of 72.6% was achieved for analyses 31 and 3231 . All results are listed in Table 5.2.3-1 and in Section 4.4.1. Individual class classification accuracies for all analyses are listed in Appendix C-2.

The effect of a 30 degrees clockwise rotation on global classification accuracy was illustrated in analysis 31*r. The

164

results from analyses 31 and 31*r are listed in Table 5.2.3-2 and in Section 4.4.3.

TABLE 5.2.3-2

Global Classification Accuracy for Analyses 31, 31*r over Image Forest

---

Analysis	Global Classification Accuracy Test Data
31 (31x31/16q)	72.6%
31*r (31x31/16q)	65.9%
(30 degree clockwise rotation)	

Although a difference exists in classification accuracies between analyses 31 and 31*r this exercise served to demonstrate the relative insensitivity of all textural features to image rotation. Section 2.3 provides further explanation.

The process of final optimal feature space derivation was tested for analyses 21, 25 and 31 by performing classification exercises of similar training and test data over subsets of features consisting of equal number of features derived from their initial optimal feature spaces. Subsets #1 features were derived from discriminant analysis while subsets #2 features were selected at random from those remaining. Results are listed in Table 5.2.3-3 and in Section 4.4.4.

TABLE 5.2.3-3

Global Classification Accuracy for Analyses 21, 25, 31 Subsets' #1 and #2 over Image Forest

	Global Classification Accuracy			
	Training Data		Test Data	
Analysis	Subset #1	Subset #2	Subset #1	Subset#2
21 (21x21/16q)	100.0%	98.0%	78.6%	71.3%
25 (25x25/16q)	100.0%	99.3%	78.0%	75.2%
31 (31x31/16q)	100.0%	97.3%	73.3%	62.0%

Global classification accuracies of image training and test data in final optimal feature spaces for subsets #1 were proven consistently superior to those of subsets #2. Small differences between classification accuracies of subsets are attributed to the ability of features to constitute feature spaces within which classes are amply discriminated. The superior classification accuracies for subsets #1 over subsets #2 validated the technique of final optimal feature selection by discriminant analysis from initial optimal spaces for all analyses.

Initial optimal features brightness, average, variance, angsecaver*, corraver*, corrang, varrang*, indiffrang, sumenrang*, entrang, meascoraverb, maxprang, ent2, and gradnt were common to 3 or more of the five analyses 21, 25, 31, 3221 and 3231. All initial optimal features annotated by an asterisk * were initial optimal features from analyses 162, 164, 322, and 324 over image Urban and from analyses 17, 21, 25, 21*3, 25*5, and 3225 over image Field. These initial optimal features were also final optimal features for 3 or more of the five analyses with exception to features brightness and gradnt. Appendix C-1 lists all final and initial optimal features for analyses over image Forest. The reoccurrence of features in multiple analyses over all three images indicates that other images displaying distinct textures could be accurately classified within feature spaces containing the latter entities.

6.0 Conclusion and Recommendations

Textural features were used to construct final optimal feature spaces and to classify monochrome image pixels for all analyses. The ability of textural features to provide for an optimal environment for classification was function of the analyst's ability to establish parameters which would allow for maximum distinction between class textures.

Textural features were very appropriate for classification of monochrome imagery provided that distinct textures exist over all classes. It is suggested that they be used with spectral features when multispectral image data is available to increase on classification accuracy. Classification accuracy was influenced by textural feature parameter settings, size of image class and sampling windows, proper training of classifiers, and the values of thresholds for feature selection exercises. High classification accuracy was possible if class textures were unique amongst themselves, few in number, and included large numbers of representative training pixels. Classification accuracy was also influenced by the ability of the analyst to properly carry out tasks for supervised parametric classification.

Low classification accuracies for several classes were attributed to improper textural feature parameter setting and selection of

unrepresentative class training pixels. Results would be greatly increased by grouping textures into common classes (i.e. all textures belonging to a ground cover class entitled "desert" would score accurate classification if desert pixels were classified into any of its textures). Classification accuracies were not recorded in this manner as all textures were assumed independent from one another and characteristic of separate classes. Reoccurrence of final optimal features over all analyses also permits the analyst to select these quantities for near-optimal classification of monochrome imagery.

It is difficult to establish universal trends in classification accuracy for different feature parameter settings. This is because classification accuracies depend upon the textures being analysed and vary from image to image. For all analyses in this study, it was shown that changes in parameter settings resulted in selection of different final optimal features and for changing classification accuracy over all classes. It is suggested that future investigation be directed towards optimizing parameter settings over all textures to provide an optimal environment for their classification. For all analyses over all images in this study, Appendix F provides a view of the effects on classification accuracy from changing parameter settings.

Classification accuracies from all analyses were comparable to those obtained by other researchers referenced in this study (considering that several images used in this study contained increased number of classes and classes of smaller size. A different approach to feature selection was also implemented).

Initial optimal feature spaces were derived from analysis of pooled class covariance matrices and final optimal features were derived by execution of multivariate discriminant analysis from initial optimal feature spaces. Final optimal feature spaces were sub-optimal in nature due to the sub-optimal feature selection technique used to derive initial final optimal features. By limiting classification accuracy in this manner significant increases in feature selection efficiency were achieved. Attention is to be directed towards optimizing initial optimal feature selection (i.e. by deriving a more representative pooled covariance matrix for all classes) without decreasing the efficiency of this process.

Pursuant to this study, it is recommended that the source code be reformatted to render all processes more effective. This includes removing all run length type features from analysis and scanning image pixels from within image class windows down consecutive columns. Use of the Moore-Penrose generalized inverse over individual class covariance matrices for derivation of final optimal

features and for classification of image data was also to be analysed in order to provide for optimal process accuracy.

7.0 <u>REFERENCES</u>

1. C. Gonzalez and P. Wintz, "Digital Image Processing", Second Edition, Addison-Wesley Publishing Co., California, 1987, pp. 13-59.

2. H. Swain and S. Davis, "Remote Sensing: The Quantitative Approach", McGraw-Hill, Inc., New York, 1978, pp. 138.

3. K. S. Fu, "Sequential Methods in Pattern Recognition and Machine Learning", Academic Press, New York, 1968.

4. T. Pavlidis, "Structural Pattern Recognition", Springer Verlag, 1977.

5. Reference 2, pp. 145-146.

6. P. A. Devijver and J. Kittler, "Pattern Recognition Theory and Applications", Springer-Verlag, New York, 1986, pp. 3.

7. K. S. Fu, "A Step Towards Unification of Syntactic and Statistical Pattern Recognition ", <u>IEEE Trans. Pattern Analysis and Machine Intelligence</u>, Vol. 8, 1986, pp. 398-404.

8. J.A. Richards, "Remote Sensing Digital Image Analysis, An Introduction", Springer-Verlag, New York, 1986, pp. 173.

9. E. L. Maxwell, "Multivariate Systems Analysis of Multispectral Imagery", <u>Photogrammetric Engineering and Remote Sensing</u>, Vol. 42, No. 9, September 1976, pp. 1173-1186.

10. Reference 8, pp. 191-205.

11. F. R. Fromm and R. A. Northhouse, "CLASS: A Non-Parametric Clustering Algorithm", <u>Pattern Recognition</u>, 8, 1976, pp. 107-114.

12. E. P. F. Kan, "ISODATA: Thresholds for Splitting Clusters", Technical Report 640-TR-058. Lockheed Electronics Co. Inc., HASD, Houston, Texas, 1972.

13. J. Bryant, "On the Clustering of Multi-dimensional Pictorial Data", Pattern Recognition, 1979, pp.115-125.

14. R. D. Duda and P.E. Hart, "Pattern Classification and Scene Analysis", New York, John Wiley and Sons, 1973.

15. Reference 6, pp. 3-4.

16. N. J. Nilsson, "Learning Machines", McGraw and Hill, New York, 1965.

17. Reference 2, pp. 146.

18. Reference 6, pp. 3.

19. Reference 2, pp. 151.

20. K. Fukunaga and T. E. Flick, "A Test of the Gaussian-ness of a Data Set using clustering", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 8, 1986, pp. 240-247.

21. R. A. Schowengerdt, "Techniques for Image Processing & Classification in Remote Sensing", Academic Press, Inc., 1983, pp 143-145.

22. R. Harris, "Satellite Remote Sensing - An Introduction", Routledge & Kegan Paul, New York, 1987, pp. 90.

23. Reference 2, pp. 151.

24. Reference 21, pp. 39.

25. Reference 8, pp. 264-266.

26. Reference 16.

27. Reference 14.

28. R. A. Schowengerdt, "Techniques for Image Processing & Classification in Remote Sensing", Academic Press, Inc., 1983.

29. H. C. Andrews, "Introduction to Mathematical Techniques in Pattern Recognition, Wiley, New York, 1972.

30. P. E. Anuta, "Geometric Correction of ERTS1 Digital Multispectral Scanner Data", LARS Information Note 103073, Purdue University, Laboratory for Applications of Remote Sensing, 1973.

31. J. D. Fahnestock and R. A. Schowengerdt, "Spatially-Variant Contrast Enhancement Using Local Range Modification", Optical Engineering, Vol. 22, No. 3, May - June 1983.

32. R. C. Gonzalez and P. Wintz, Digital Image Processing, Reading, Mass., Addison-Wesley, 1977.

33. J. S. Weszka, C. R. Dyer and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-6, No. 4, April 1976, pp. 269-285.

34. E. B. Troy, E. S. Deutsch and A. Rosenfeld, "Gray-Level Manipulation Experiments for Texture Analysis", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-3, No. 1, January 1973, pp. 91-98.

35. T. M. Lillesand and R. W. Kiefer, "Remote Sensing and Image Interpretation", John Wiley and Sons, New York, 1979, pp. 610-705.

36. Reference 2, pp. 159-163.

37. Reference 2, pp. 163-174.

38. A. G. Wacker, "The Minimum Distance Approach to Classification", Ph. D. Thesis, Purdue University, West Lafayette, 1971.

39. Reference 2, pp. 164-174.

40. C. Chen, "Statistical Pattern Recognition", Hayden Book Company, Inc., Rochelle Park, New York, 1973, pp. 59.

41. P. H. Swain, T. V. Robertson and A. G. Wacker, "Comparison of Divergence and B-Distance in Feature Selection, Information Note 020871, Lab. for Applications of Remote Sensing, Purdue Univ., West Lafayette.

42. Reference 28, pp. 44,149.

43. Reference 2, pp. 229.

44. R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, Vol SMC-3, No 6, November 1973, pp. 610-621.

45. R. L. Keltig and D. A. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects", IEEE Transactions on Geoscience Electronics, GE-4, 1, pp. 19-26.

46. R. M. Haralick, "Statistical and Structural Approaches to Texture", Proceedings of the IEEE, Vol. 67, No. 5, May 1979, pp. 786-804.

47. Reference 33.

48. R. M. Haralick, "A Texture-Context Feature Extraction Algorithm for Remotely Sensed Imagery", Proc. 1971 IEEE Decision and Control Conf., (Gainsville, FL), December 15-17, 1971, pp. 650-657.

49. Reference 44.

50. R. M. Haralick and K. S. Shanmugam, "Combined Spectral and Spatial Processing of ERTS Imagery Data", Remote Sensing of the Environment, 1974, 3:3-13.

51. Reference 44.

52. R. Conners, "Towards a Set of Statistical Features which Measure Visually Perceivable Qualities of Texture", CH1428-2/79/0000-0382 IEEE, 1979, pp. 382-389.

53. Reference 33.

54. M. M. Galloway, "Texture Classification using Gray Level Run Lengths", Comp. Graphics and Image Proc., Vol. 4, June 1975, pp. 172-179.

55. Reference 33.

56. R. Sutton and E. Hall, "Texture Measures for Automatic Classification of Pulmonary Disease", IEEE Transactions on Computers, Vol. C-21, No. 7, July 1972, pp. 667-676.

57. A. Rosenfeld, "A Note on Automatic Detection of Texture Gradients", IEEE Transactions on Computers, October 1971, pp. 988-991.

58. A. Rosenfeld and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis", IEEE Transactions on Computers, 1971, C-20:, pp. 562-569.

59. Reference 14.

60. Reference 33.

61. S. Hsu, "Texture-Tone Analysis for Automated Landuse Mapping", Photogram. Eng. Remote Sensing, 44:1393-1404.

62. J. R. Irons and G. Peterson, "Texture Transforms of Remote Sensing Data", Remote Sensing of the Environment, November 1981, pp. 359-370.

63. J. E. Estes, E. J. Hajic and L. R. Tinney, "Fundamentals of Image Analysis of Visible and Thermal Infrared Data", Manual of Remote Sensing, Volume 1, Second Edition, D. S. Simmonett and F. T. Ulaby, eds., American Society of Photogrammetry, Virginia, 1983, pp. 1044.

64. Reference 8, pp. 206.

65. G. Xuan, "A New Feature Selection Method Based on the Mahalanobis Distance", CH2046-1/84/0000/0131, IEEE, 1984, pp. 131-133.

66. S. Morgera and L. Datta, "Optimal Feature Selection", Parts 1 and 2, CH2046-1/84/0000/0134, IEEE, 1984, pp. 134-141.

67. Reference 40, pp. 57-62.

68. Reference 8, pp. 210-211.

69. J. Schott, E. Krauss, C. Salvaggio, "Optimum Spectral Band Selection", DIRS Report #88/89-54-117, July 30 1988, pp. 30.

70. E. P. Crist and R. J. Kauth, "The Tassled Cap De-Mystified", Photogrammetric Engineering and Remote Sensing, 52, 1, 1986, pp. 81-86.

71. Reference 12, pp. 214-218.

72. P. O. Peitgen and P. H. Richter, "The Beauty of Fractals", Springer-Verlag, New-York, 1986, pp. 199.

73. Reference 28, pp. 54, 177.

74. Reference 21, pp. 184.

75. IMSL Problem-Solving Software Systems Math Library, Fortran Subroutines for Mathematical Applications, Volume 1, Chapters 1-2, Version 1-0, April 1987, MAL-USM-UNBND-1.0, pp. 286,288.

76. R. A. Johnson, D. W. Wichern, "Applied Multivariate Statistical Analysis", Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982, pp. 151.

77. Reference 21, pp. 147.

8.0 APPENDICES

Appendix A

A) <u>Tonal Features</u> : Tonal features used for classification of digital image data consist of:

i) *brightness* or *gray-level counts* of individual image pixels, or

ii) *means, standard deviations* and *range* of gray-level counts over image segments.

B) <u>Textural Features</u> : Textural features used for classification of digital image data consist of:

i) *Cooccurrence Matrix Textural Features* : The following fourteen textural features are extracted from gray-tone cooccurrence matrices. For a chosen image sampling distance and image window size, four angular gray-tone cooccurrence matrices can be derived for each feature. The mean and range of each of the fourteen textural feature averaged over all four directions consist of inputs

A-1

to the classifier and twenty-eight features are formed.

Notation :   - p(i,j) : (i,j)th entry in a normalized gray-tone spatial dependence matrix, = P(i,j)/R

- $p_x(i)$: ith entry in the marginal-probability matrix obtained by summing the rows of p(i,j) where:

$$p(i,j) = \sum_{j=1}^{N_g} P(i,j).$$

- Ng : number of gray levels in the quantized image.

- R : number of resolution cells in the cooccurrence matrix

$$p_y(j) : \sum_{i=1}^{N_g} p(i,j).$$

$$- p_{x+y}(k) : \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j). \qquad k = 2,3,\ldots\ldots\ldots 2N_g.$$

$$i + j = k$$

$$- p_{x-y}(k) : \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j). \qquad k = 0,1,\ldots\ldots\ldots N_g - 1$$

$$/i - j/ = k$$

1) <u>Angular Second Moment</u>: $f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \{ p(i,j)\}^2.$

2) <u>Contrast</u>: $f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}.$

$$/i-j/=n$$

3) <u>Correlation</u>: $f_3 = \dfrac{\left[\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i,j)\, p(i,j) - \mu_x \mu_y \right]}{(var\ x)(var\ y)}$

4) <u>Variance</u>: $f_4 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu)^2\, p(i,j).$

5) <u>Inverse Difference Moment</u>: $f_5 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left\{ \dfrac{p(i,j)}{(1+(i-j)^2)} \right\}$

6) <u>Sum Average</u>: $f_6 = \sum_{i=2}^{2N_g} i\, p_{x+y}(i)$.

7) <u>Sum Variance</u>: $f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2\, p_{x+y}(i)$.

8) <u>Sum Entropy</u>: $f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i)\, \log\{p_{x+y}(i)\}$.

9) <u>Entropy</u>: $f_9 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)\, \log(p(i,j))$.

10) <u>Difference Variance</u>: f_{10} = variance of p_{x-y}.

11) <u>Difference Entropy</u>: $f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i)\, \log\{p_{x-y}(i)\}$.

12), 13) Information measures of correlation:

$$f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}}.$$

$$f_{13} = (1 - \exp[-2.0(HXY2 - HXY)])^{1/2}.$$

where:

$$HXY = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j)).$$

$$HXY1 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log\{p_x(i)p_y(j)\}.$$

$$HXY2 = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_x(i)p_y(j) \log\{p_x(i)p_y(j)\}$$

14) Maximal Correlation Coefficient:

$$f_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}.$$

$$Q(i,j) = \sum_{k} \left\{ \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)} \right\}.$$

A- 5

Another approach to defining features of this class is to use matrices based on pairs of *average gray levels*, taken over neighbourhoods whose centers are ∂ distance apart.

ii) *Gray Level Difference Statistics* : Let $f_m(x,y)$ be the average gray-level of the image in a square region of side m+1 approximately centered at (x,y), or the gray-level of the same square region centered at (x,y). The choice is left to the analyst. Features are the differences of the latter entities for pairs of horizontally, vertically, or diagonally adjacent regions as follows:

1) $/f_m(x-m, y-\frac{[m+1]}{2}) - f_m(x+1, y-\frac{[m+1]}{2})/$ (horizontal)

2) $/f_m(x-\frac{[m+1]}{2}, y-m) - f_m(x-\frac{[m+1]}{2}, y+1)/$ (vertical)

3) $/f_m(x-m, y-m) - f_m(x, y)/$ and

$\qquad\qquad\qquad\qquad\qquad\qquad$ (diagonals).

$/f_m(x-m, y) - f_m(x, y-m)/$

: additional features derived from absolute differences between pairs of gray levels or of average gray levels. For for a given displacement ∂ = $(\Delta x, \Delta y)$ we let $f_\partial(x,y)$ = $/f(x,y) - f(x + \Delta x, y + \Delta y)/$ and p_∂ the probability density of $f_\partial(x,y)$ we derive the following features which measure image coarseness:

1) <u>Contrast</u>: $\text{CON} = \sum_i i^2 p_\partial(i)$.

2) <u>Angular 2nd Moment</u>: $\text{ASM} = \sum_i p_\partial(i)^2$.

3) <u>Entropy</u>: $\text{ENT} = -\sum_i p_\partial(i) \log p_\partial(i)$.

4) <u>Mean</u>: $M = (1/m) \left(\sum_i i\, p_\partial(i) \right)$. (for m gray levels).

5) <u>Inverse Difference Moment</u>: $\text{IDM} = \sum_i p_\partial(i) / (i^2 + 1)$.

A- 7

Another approach to defining features of this class is to use p_∂ based on differences between pairs of *average gray levels*, taken over neighbourhoods whose centers are ∂ distance apart.

iii) *Gray Level Run Statistics* : Let $p(i,j)$ be the number of runs of length j, in directions $\beta = 0, 45, 90, 135$ degrees, consisting of points whose gray-levels have value i. Let Nr be the number of runs and Ng be the number of gray-levels. The features which make up this type of measurement are listed as follows :

1) <u>Long Runs Emphasis</u> : <u>for any gray-level i</u>

$$LRE = \sum_{j=1}^{Nr} j^2 p(i,j) \bigg/ \sum_{j=1}^{Nr} p(i,j).$$

2) <u>Short Runs Emphasis</u> : <u>for any gray-level i</u>

$$SRE = \sum_{j=1}^{Nr} \left(\frac{p(i,j)}{j^2} \right) \bigg/ \sum_{j=1}^{Nr} p(i,j).$$

3) <u>Gray Level Distribution</u> :

$$GLD = \sum_{i=1}^{Ng} \left(\sum_{j=1}^{Nr} p(i,j) \right)^2 \bigg/ \sum_{j=1}^{Nr} p(i,j)$$

4) <u>Run Length Distribution</u> : $RLD = \sum_{j=1}^{Nr} \left(\sum_{i=1}^{Ng} p(i,j) \right)^2 / \sum_{j=1}^{Nr} p(i,j).$

5) <u>Run Percentage</u> : $RPC = \left(\sum_{j=1}^{Nr} \frac{p(i,j)}{N^2} \right).$ (N is the number of
<u>for any gray-level i</u> points in the image).

6) <u>Long Runs Emphasis</u> : $LRM = \sum_{i=1}^{Ng} \sum_{j=1}^{Nr} j^2\, p(i,j) / \sum_{i=1}^{Ng} \sum_{j=1}^{Nr} p(i,j).$
<u>moments</u>

7) <u>Short Runs Emphasis</u> : $SRM = \sum_{i=1}^{Ng} \sum_{j=1}^{Nr} \frac{p(i,j)}{j^2} / \sum_{i=1}^{Ng} \sum_{j=1}^{Nr} p(i,j).$
<u>moments</u>

8) <u>Gray Level Non-</u> : $GLN = \sum_{i=1}^{Ng} \left(\sum_{j=1}^{Nr} p(i,j) \right)^2 / \sum_{j=1}^{Nr} p(i,j).$
<u>Uniformity</u>

A- 9

9) **Run Length Non-** : RLD $= \sum\limits_{j=1}^{Nr} \left(\sum\limits_{i=1}^{Ng} p(i,j) \right)^2 / \sum\limits_{i=1}^{Ng} \sum\limits_{j=1}^{Nr} p(i,j)$

Uniformity

10) **Fraction of Image** : FIR $= \sum\limits_{i=1}^{Ng} \sum\limits_{j=1}^{Nr} p(i,j) \bigg/ \sum\limits_{i=1}^{Ng} \sum\limits_{j=1}^{Nr} p(i,j).$

in Runs

iv) *Composite Features* : from a given set of features, a set of composite features is derived (1) by taking the *mean* s, *ranges* and *standard deviations* over all sampling directions for each image sampling distance for each feature, and (2) by taking the *means, ranges* and *standard deviations* over all sampling orientations, for each sampling direction for each feature.

v) *Textural Edgeness* : Features extracted from image gradients provide a measure of the degree of edgeness in image areas as follows:

- For every sampling distance d and resolution cell
I defined over neighbourhood N the gradient may be
expressed as follows:

$$G(d) = \sum_{[(i,j)\,=\,N]} \left\{ /I(i,j) - I(i+d,j)/ + /I(i,j) - I(i-d,j)/ + /I(i,j) - I(i,j+d)/ + /I(i,j) - I(i,j-d)/ \right\}$$

(The quick Robert's, Sobel and Prewitt gradient
filters may also be applied upon image data)

- following computation of a gradient for a desired
subimage and possible subsequent gray-level
thresholding the *average value of the gradient*
$A_r(x,y)$ over a neighbourhood of radius r centered at
(x,y) is determined as follows:

$$A_r(x-r,y) - A_r(x+r,y) = H_r(x,y) \quad (\text{horizontal}$$
rate of change of average brightness).

$$A_r(x,y-r) - A_r(x,y+r) = V_r(x,y) \quad (\text{vertical rate of}$$
change of average brightness).

and, $(H_r^2 + V_r^2)^{1/2}$ (magnitude of maximum
rate of change).

$$\tan^{-1}(V_r/H_r) \qquad (\text{ direction of maximum rate of change }).$$

vi) *Texture Transform Features* : From each channel of multispectral image data are derived texture transforms using user-specified windows of a given dimension. Each texture transform represents the spatial distribution of gray-levels in and around each pixel of the original image in a given spectral channel of the image data. For single channel image data, the features which are extracted from the texture transform include the *mean, standard deviation, skewness and kurtosis* of the gray-levels within the windows, *mean gray-level differences* between nearest-neighbor pixels and the *mean area* above and below tonal thresholds. For p-channel multispectral image data the following features are extracted from user-specified windows which take into account the additional dimensionality of the image data:

1) <u>Mean (MNL)</u> : $\sum x_{ij}/n$ (x_{ij} is gray level vector norm length for image pixel(i,j)).

2) <u>Variance (VNL)</u> :
$$\frac{\sum (x_{ij} - MNL)^2}{n - 1}$$

3) <u>Skewness (SKEW)</u> :
$$\frac{|\sum (x_{ij} - MNL)^3|}{(n - 1)(VNL)^{3/2}}$$

4) <u>Kurtosis (KURT)</u> :
$$\frac{\sum (x_{ij} - MNL)^4}{(n - 1)(VNL)^2}$$

5) <u>Range (RNL)</u> : $max(x_{ij}) - min(x_{ij})$

6) <u>Pearson's 2nd Coefficient of Skewness (PSKEW)</u> :

$$\frac{|MNL - x_m|}{(VNL)^{1/2}}$$ (where x_m = median norm length in a window).

7) <u>Absolute Value of Mean Norm Length Differences (MDIF)</u> :

$$\frac{|\sum (x_{ij} - x_c)|}{n - 1}$$ (where x_c = norm length of the gray level vector representing a window's center pixel).

8) <u>Mean of Squared Norm Length Differences (MSQ)</u> :

$$\frac{\sum (x_{ij} - x_c)^2}{n - 1}$$

9) <u>Maximum of Squared Norm Length Differences (MAXSQ)</u> :

$$\max(x_{ij} - x_c)^2$$

10) <u>Mean Euclidean Distance (MEUC)</u> :

$$\frac{\sum \left[\sum_f (x_{cf} - x_{ijf})^2 \right]^{1/2}}{n - 1}$$

(where x_{ijf} = gray tone for spectral channel f, pixel (i,j) of a multispectral image), and

(x_{cf} = gray tone for spectral channel f of a window's center pixel).

11) <u>Maximum Euclidean Distance (MAXEUC)</u> :

$$\max \left[\sum_f (x_{cf} - x_{ijf})^2 \right]^{1/2}$$

A- 14

Appendix A-1

Textural Features Utilized for all Analyses on Images Urban, Field, and Forest

‑‑

	Image:	Urban./Field./Forest.	
A-Cooccurrence Type Features			
1- angular second moment average (*angsecaver*)	Y	Y	Y
2- angular second moment range (*angsecrang*)	Y	Y	Y
3- contrast average (*contraver*)	Y	Y	Y
4- contrast range (*contrang*)	Y	Y	Y
5- correlation average (*corraver*)	Y	Y	Y
6- correlation range (*corrang*)	Y	Y	Y
7- variance average (*varraver*)	Y	Y	Y
8- variance range (*varrang*)	Y	Y	Y
9- inverse difference moments average (*indiffaver*)	Y	Y	Y
10- inverse difference moments range (*indiffrang*)	Y	Y	Y
11- sum average average (*sumaver*)	Y	Y	Y
12- sum average range (*sumavrang*)	Y	Y	Y

	Urban	Field	Forest
13- sum variance average (*sumvaver*)	Y	Y	Y
14- sum variance range (*sumvarang*)	Y	Y	Y
15- sum entropy average (*sumenaver*)	Y	Y	Y
16- sum entropy range (*sumenrang*)	Y	Y	Y
17- entropy average (*entaver*)	Y	Y	Y
18- entropy range (*entrang*)	Y	Y	Y
19- difference entropy average (*diffenaver*)	Y	Y	Y
20- difference entropy range (*diffenrang*)	Y	Y	Y
21- information measure of correlation A av (*meascoravera*)	Y	Y	Y
22- information measure of correlation A range (*meascoranga*)	Y	Y	Y
23- information measure of correlation B av (*meascoraverb*)	Y	Y	Y
24- information measure of correlation B range (*meascorangb*)	Y	Y	Y
25- difference variance average (*diffvaraver*)	Y	Y	Y
26- difference variance range (*dvarg*)	Y	Y	Y
27- maximum probability average (*maxpaver*)	Y	Y	Y
28- maximum probability range (*maxprang*)	Y	Y	Y

B - First order Statistics

	Urban	Field	Forest
29- gradient (*gradnt*)	Y	Y	Y

	Urban	Field	Forest
30- brightness (*bright*)		Y	Y
31- average brightness (*moyenne*)	Y	Y	Y
32- variance (*ecart*)	Y	Y	Y

C - Run Length Statistics

	Urban	Field	Forest
33- short run emphasis inverse moment av. (*sreimaver*)	Y	Y	Y
34- short run emphasis inverse moment range (*sreimrang*)	Y	Y	Y
35- long run emphasis inverse moment av. (*lreimaver*)	Y	Y	Y
36- long run emphasis inverse moment range (*lreimrang*)	Y	Y	Y
37- grey level non-uniformity average (*glnaver*)	Y	Y	Y
38- grey level non-uniformity range (*glnrang*)	Y	Y	Y
39- run length non-uniformity average (*rlnaver*)	Y	Y	Y
40- run length non-uniformity range (*rlnrang*)	Y	Y	Y
41- fraction of image in runs average (*firraver*)	Y	Y	Y
42- fraction of image in runs range (*firrang*)	Y	Y	Y

D - Grey Level Difference Statistics

	Urban	Field	Forest
43- contrast (*contrast2*)	Y	Y	Y
44- angular second moment (*angsec2*)	Y	Y	Y

A-1- 3

	Urban	Field	Forest
45- entropy (*ent2*)	Y	Y	Y
46- mean (*mean2*)	Y	Y	Y

_____

note : av. = average,
     Y = feature was used in all analyses for image under
        consideration

APPENDIX A-2

Summary of Initial and Final Optimal Feature Spaces for Analyses 162, 164, 322, 324 for Image Urban

Analysis : Features	162	164	322	324
angsecaver	*/@	*/@	*/@	*/@
angsecrang	*	*/@	*	*/@
contraver	*/@	*	*/@	*
corraver	*	*	*/@	*/@
corrang				*
varraver	*/@	*/@	*/@	*/@
varrang	*	*	*	*/@
indiffrang		*/@	*	*/@
sumenrang	*/@	*/@	*	*/@
entrang		*/@		*/@
diffenrang	*/@	*/@	*/@	*/@
meascoranga	*/@	*	*/@	*/@
meascoraverb		*		*/@
meascorangb	*	*		*/@
sreimrang		*		*/@
lreimrang	*		*/@	
ent2			*/@	
glnaver		*		*
glnrang		*		
rlnaver		*		
rlnrang	*			
gradnt				*

note : * = feature was derived for initial optimal feature space

@ = feature was derived for final optimal feature space

Percentage Classification Accuracy of Training and Test
Data for Analyses 162, 164, 322, 324
for Image Urban
(individual class covariance matrices)

Analysis:	162		#164		#322		#324	
Class	Train	Test	Train	Test	Train	Test	Train	Test
1	96.8	60..0	97.1	33.3	100.0	76.6	100.0	50.0
2	97.2	43.0	100.0	33.3	100.0	60.0	100.0	20.0
3	97.1	30.0	100.0	10.0	100.0	43.3	0.0	0.0
4	89.6	86.6	95.4	83.3	91.4	93.3	98.0	96.6
5	80.0	83.3	92.6	90.0	80.6	83.3	97.2	83.3
6	77.1	43.3	92.8	33.3	77.1	40.0	75.9	26.6
7	83.9	73.3	89.0	50.0	75.2	66.6	86.4	56.6
8	63.5	60.0	72.1	70.0	57.3	63.3	81._	63.3
9	56.4	36.6	75.3	36.6	92.3	70.0	97.4	70.0
10	90.4	66.6	95.5	60.0	90.4	56.6	100.0	43.3
11	85.9	66.6	91.8	66.6	71.3	50.0	99.1	33.3

APPENDIX A-4

Percentage Classification Accuracy of Training Data for Analyses (162, 261), (164, 461), (322, 223), (324, 423) for Image Urban

Analysis :	162	261	164	461
Class				
1	96.8	96.8	97.1	100.0
2	97.2	97.2	100.0	100.0
3	97.1	97.1	100.0	100.0
4	89.6	85.0	95.4	97.5
5	80.0	74.5	92.6	79.2
6	77.1	80.1	92.8	96.1
7	83.9	87.3	89.0	58.1
8	63.5	60.7	72.1	56.3
9	56.4	94.8	75.3	48.1
10	90.4	88.0	95.5	86.6
11	85.9	4.0	91.8	95.1

Analysis :	322	223	324	423
Class				
1	100.0	100.0	100.0	100.0
2	100.0	98.6	100.0	100.0
3	100.0	100.0	0.0	100.0
4	91.4	90.8	98.0	99.0
5	80.6	80.9	97.2	98.6
6	77.1	77.5	75.9	93.5
7	75.2	77.7	86.4	93.9

	322	223	324	423
8	57.3	53.4	81.3	79.2
9	92.3	80.3	97.4	97.4
10	90.4	91.0	100.0	100.0
11	71.3	90.0	99.1	95.9

APPENDIX A-5

Percentage Classification Accuracy of Test Data for Analyses (162, 261), (164, 461), (322, 223), (324, 423) for Image Urban

Analysis :	162	261	164	461
Class				
1	60.0	73.3	33.3	43.3
2	43.0	46.6	33.3	20.0
3	30.0	40.0	10.0	16.6
4	86.6	93.3	83.3	90.0
5	83.3	73.3	90.0	50.0
6	43.3	56.6	33.3	50.0
7	73.3	76.6	50.0	50.0
8	60.0	56.6	70.0	36.6
9	36.6	86.6	36.6	13.3
10	66.6	66.6	60.0	60.0
11	66.6	6.6	66.6	66.6

Analysis :	322	223	324	423
Class				
1	76.6	76.6	50.0	63.3
2	60.0	60.0	20.0	43.3
3	43.3	43.3	0.0	10.0
4	93.3	86.6	96.6	96.6
5	83.3	76.6	83.3	93.3
6	40.0	36.6	26.6	43.3
7	66.6	70.0	56.6	73.3

	322	223	324	423
8	63.3	56.6	63.3	70.0
9	70.0	60.0	70.0	70.0
10	56.6	70.0	43.3	53.3
1 1	50.0	66.6	33.3	23.3

APPENDIX A-6

Percentage Classification Accuracy of Training Data
for Analyses 162, 164, 322, 324
for Image Urban
(pooled class covariance matrix)

Analysis :	162	164	322	324
Class				
1	53.9	91.4	57.1	22.9
2	38.8	55.0	52.8	5.0
3	60.0	33.3	54.3	0.0
4	74.4	94.5	79.0	91.0
5	68.8	75.9	70 .6	75.9
6	3.9	51.3	37.1	72.7
7	68.7	84.7	72.4	60.3
8	57.4	66.5	55.9	76.0
9	45.3	75.3	69.2	2.6
10	54.4	85.5	56.6	66.6
11	75.0	50.8	64.1	99.2

Summary of Individual Feature Subsets for Analyses 162, 164, 322 for Image Urban

Analyses :	162		164		322	
Subsets :	#1	#2	#1	#2	#1	#2
angsecaver		*	*		*	
angsecrang		*	*			*
contraver	*			*	*	
corraver	*			*	*	
varraver	*		*		*	
varrang		*		*		*
indiffrang			*			*
sumenrang	*		*			*
entrang				*		
diffenrang	*		*			*
meascoranga		*	*			*
meascoraverb				*		
meascorangb				*		
sreimrang			*			
lreimrang		*			*	
ent2					*	
glnaver				*		
glnrang				*		

note :   * = feature was selected within appropriate analysis subset

Percentage Classification Accuracy of Training and Test Data for Analysis Subsets 162, 164, 322 for Image Urban
(individual class covariance matrices)

Analyses:	162				164			
Subsets:	#1		#2		#1		#2	
Class	Train	Test	Train	Test	Train	Test	Train	Test
1	93.6	80..0	87.3	40.0	100.0	33.3	100.0	36..6
2	97.2	46.6	31.9	30.0	100.0	30.0	97.5	20.0
3	97.1	30.0	74.3	50.0	100.0	3.0	100.0	30.0
4	85.9	96.6	87.3	96.6	95.9	86.6	96.6	96.6
5	76.4	66.6	40.3	46.6	95.8	90.0	97.2	90.0
6	79.3	46.6	11.2	6.0	90.2	30.0	51.3	16.6
7	85.1	83.3	9.9	3.0	89.1	53.3	56.7	53.3
8	59.1	63.3	2.0	3.0	74.6	63.3	75.7	76.6
9	58.1	33.3	29.1	23.3	74.0	36.6	74.0	26.6
10	91.1	66.6	23.5	20.0	93.3	50.0	92.2	50.0
11	85.9	73.3	10.4	6.0	93.4	66.6	22.1	3.0

Analysis : **3 2 2**

Class	#1 Train	#1 Test	#2 Train	#2 Test
1	100.0	76.6	42.8	26.6
2	100.0	63.3	56.4	36.6
3	100.0	46.6	82.9	66.6
4	91.5	96.6	72.1	73.3
5	80.9	86.6	46.4	56.6
6	76.7	36.6	14.2	10.0
7	77.1	70.0	14.2	20.0
8	57.8	60.0	11.5	16.6
9	94.9	73.3	64.1	60.0
10	90.4	56.6	21.3	10.0
11	50.0	43.3	61.4	40.0

Appendix A-9

Best Combinations of Features for Analyses 162, 164, 322, 324 for Image Urban

---

Analysis 162 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f4, f5

best 3 = f3, f4, f5

best 4 = f2, f3, f4, f5

best 5 = f3, f4, f5, f7, f8

best 6 = f1, f2, f3, f5, f6, f8

best 7 = f2, f3, f5, f6, f7, f8, f9

best 8 = f2, f3, f5, f7, f8, f9, f10, f11

best 9 = f1, f2, f3, f6, f7, f8, f9, f10, f11

best 10 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10

best 11 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11

where :

Co-occurrence Type Features

      f1 = angular second moment average

      f2 = angular second moment range

      f3 = contrast average

      f4 = correlation average

      f5 = variance average

      f6 = variance range

      f7 = sum entropy range

      f8 = difference entropy range

      f9 = measure of correlation A range

Run Length Statistics

      f10 = long run emphasis inverse moment range

      f11 = run length non-uniformity range

Analysis 164 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f4, f5

best 3 = f1, f5, f16

best 4 = f1, f2, f5, f7

best 5 = f1, f2, f5, f7, f10

best 6 = f2, f4, f5, f7, f9, f10

best 7 = f1, f2, f5, f7, f8, f9, f10

best 8 = f1, f2, f5, f7, f8, f10, f11, f14

best 9 = f1, f2, f3, f5, f7, f8, f9, f11, f14

best 10 = f1, f3, f5, f6, f7, f8, f9, f10, f11, f14

best 11 = f2, f5, f7, f8, f9, f11, f12, f13, f14, f15, f16

best 12 = f1, f3, f7, f9, f10, f11, f12, f13, f14, f15, f16, f17

best 13 = f1, f2, f3, f4, f5, f6, f7, f9, f10, f11, f12, f13, f15

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f17

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f17

best 17 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17

where :

Co-occurrence Type Features

f1 = angular second moment average

f2 = angular second moment range

f3 = contrast average

f4 = correlation average

f5 = variance average

f6 = variance range

f7 = inverse difference range

f8 = sum entropy range

f9 = entropy range

f10 = difference entropy range

f11 = measure of correlation A range

f12 = measure of correlation B average

f13 = measure of correlation B range

Run Length Statistics

f14 = short run emphasis inverse moment range

f15 = grey level non-uniformity average

f16 = grey level non-uniformity range

f17 = run length non-uniformity average

Analysis 322 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f4, f5

best 3 = f1, f4, f5

best 4 = f1, f4, f5, f12

best 5 = f1, f4, f5, f11, f12

best 6 = f1, f3, f4, f5, f11, f12

best 7 = f1, f3, f4, f5, f9, f11, f12

best 8 = f1, f3, f4, f5, f9, f10, f11, f12

best 9 = f1, f3, f4, f5, f6, f9, f10, f11, f12

best 10 = f1, f2, f3, f4, f5, f6, f9, f10, f11, f12

best 11 = f1, f2, f3, f4, f5, f6, f7, f9, f10, f11, f12

best 12 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12

where :

Co-occurrence Type Features

      f1 = angular second moment average

      f2 = angular second moment range

      f3 = contrast average

      f4 = correlation average

      f5 = variance average

      f6 = variance range

      f7 = inverse difference range

      f8 = sum entropy range

      f9 = difference entropy range

      f10 = measure of correlation A range

Run Length Statistics

      f11 = long run emphasis inverse moment range

Grey Level Difference Statistics

      f12 = entropy 2

Analysis 324 Best Combinations of Features in Final Optimal Feature Space :

best 2 : f4, f6

best 3 : f1, f4, f6

best 4 : f1, f3, f4, f6

best 5 : f1, f3, f4, f6, f8

best 6 : f1, f3, f4, f6, f8, f15

best 7 : f1, f3, f4, f6, f8, f10, f15

best 8 : f1, f3, f4, f6, f8, f10, f13, f15

best 9 : f1, f3, f4, f5, f6, f7, f8, f10, f15

best 10 : f1, f2, f3, f4, f5, f6, f7, f8, f10, f15

best 11 : f1, f3, f4, f5, f6, f7, f8, f9, f10, f12, f14

best 12 : f1, f3, f4, f5, f6, f7, f8, f9, f10, f12, f13, f15

best 13 : f1, f2, f4, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

best 14 : f1, f2, f3, f4, f5, f6, f7, f8, f9, f11, f12, f13, f14, f15

best 15 : f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

best 16 : f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16

best 17 : f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f17

where :

Co-occurrence Type Features

f1 = angular second moment average

f2 = angular second moment range

f3 = contrast average

f4 = correlation average

f5 = correlation range

f6 = variance average

f7 = variance range

f8 = inverse difference range

f9 = sum entropy range

f10 = entropy range

f11 = difference entropy range

f12 = measure of correlation A range

f13 = measure of correlation B average

f14 = measure of correlation B range

Run Length Statistics

f15 = short run emphasis inverse moment range

f16 = grey level non-uniformity average

First Order Statistics

f17 = gradient

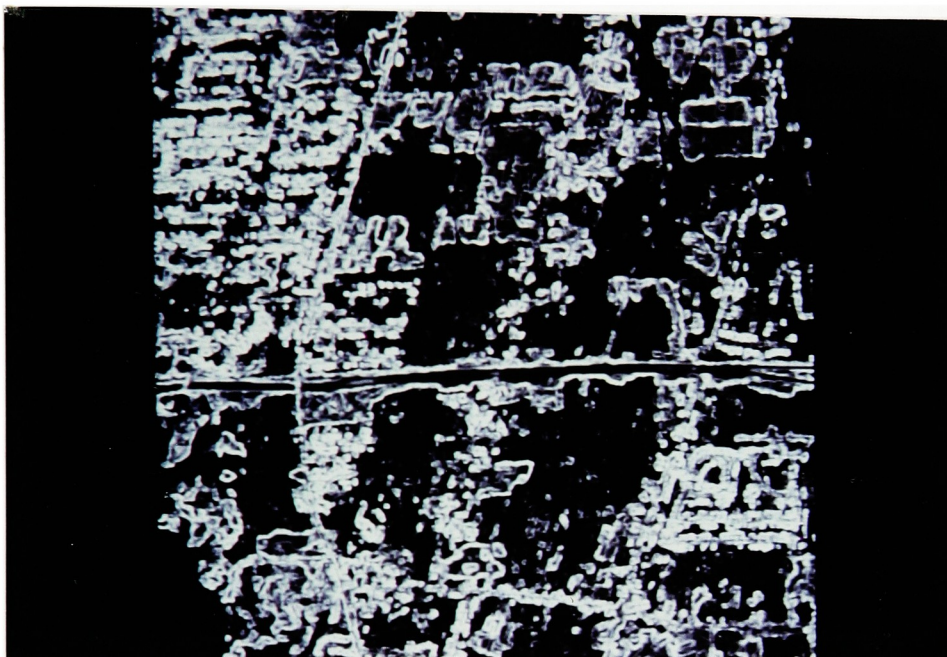Angular Second Moment Average (Angsecaver)

Contrast Average (Contraver)

Correlation Average (Corraver)
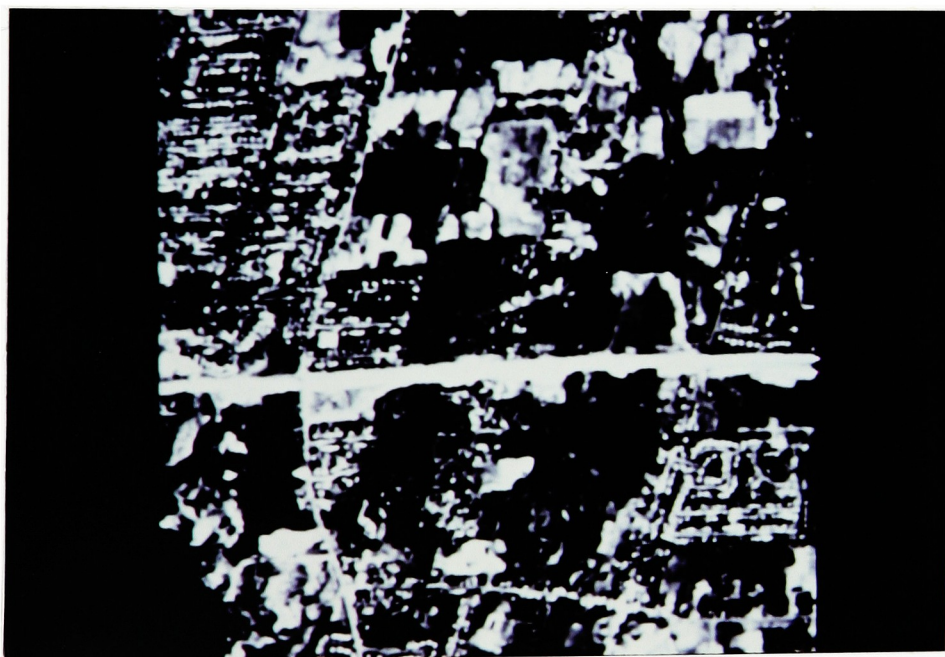
Variance Average (Varraver)
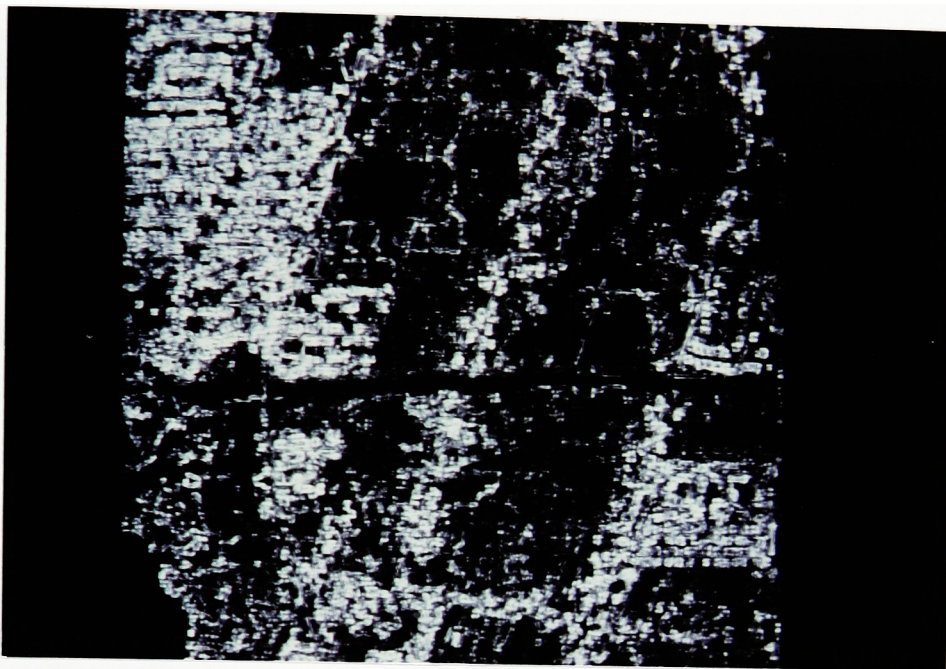
Variance   Range   (Varrang)
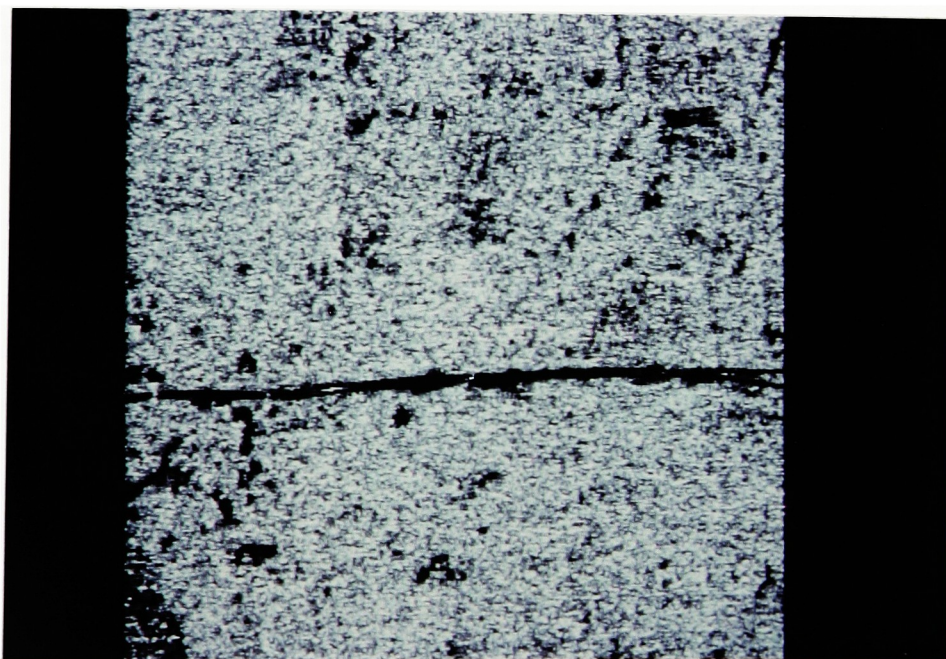
Sum   Entropy   Range   (Sumenrang)
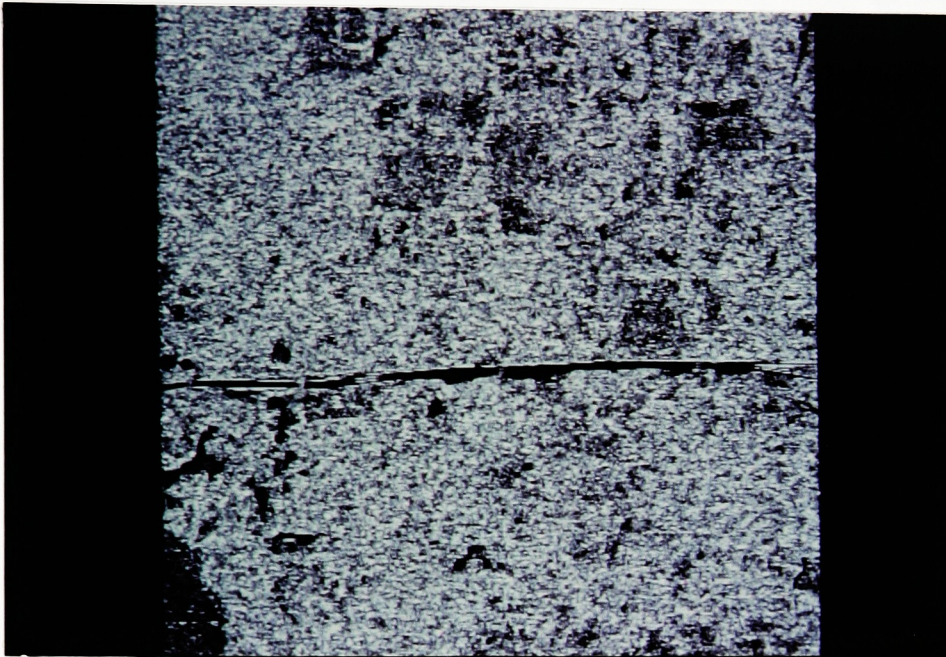
Difference Entropy Range (Diffenrang)
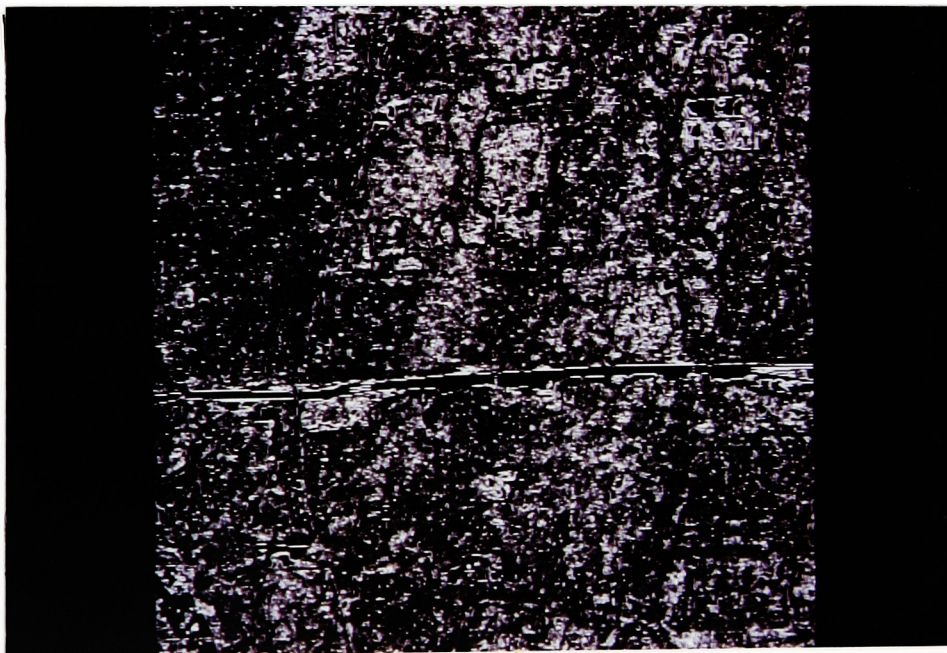
Measure of Correlation A Range (Meascoranga)
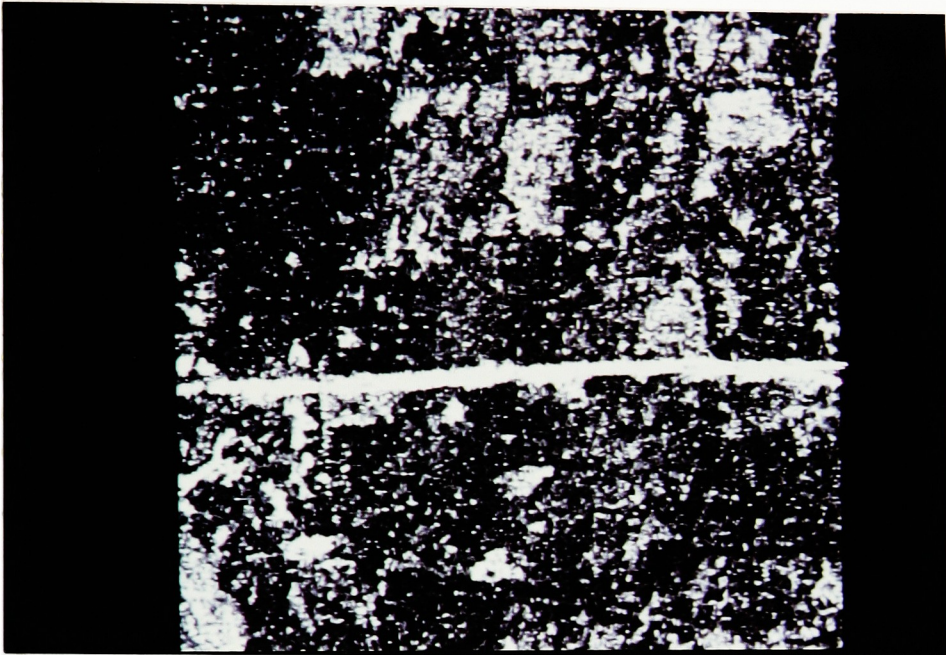
Long Run Emphasis Inverse Moment Range (Lreimrang)
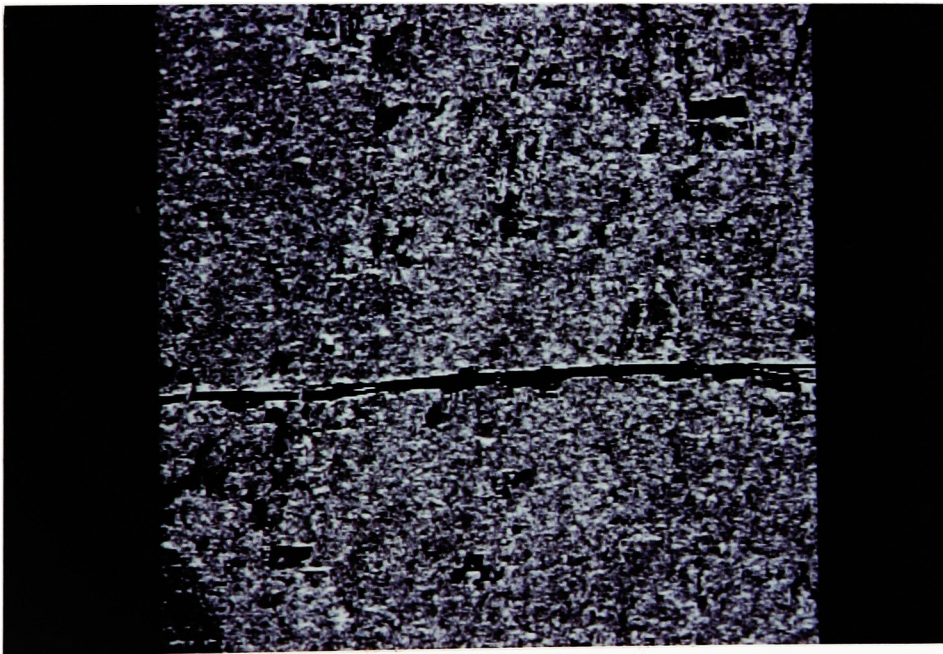
Run Length Non-Uniformity Range (Rlnrang)

APPENDIX B-1

Summary of Initial and Final Optimal Feature Spaces for Analyses 17, 21, 25, 21*3, 25*5, 3225 for Image Field

Analysis :	17	21	25	21*3	25*5	3225
Features						
brightness	*	*	*	*	*	
average	*@	*@	*@	*@	*@	*@
variance	*@	*@	*@	*@	*@	*
angsecaver	*@	*@		*@		*@
angsecrang	*@	*@	*@	*@	*@	
contraver	*@					
contrang					*@	
corraver		*@	*@	*@	*@	*@
corrang	*@	*@	*@	*@	*@	*
varrang	*@	*@	*@	*@	*	*@
indiffrang	*@	*@		*@	*@	*@
sumenrang	*@	*	*@	*@		*@
entrang				*@	*@	
diffenrang	*@	*@		*@	*@	
meascoravera						
meascoranga	*@	*@	*@	*@	*	*@
meascoraverb	*@		*@	*@	*@	
dvarg			*			*
maxprang			*@			*
sreimrang	*@	*@	*@	*@	*	*@
lreimrang						*

	17	21	25	21*3	25*5	3225
glnaver		*@				
glnrang				*@	*@	
contrast2		*@		*		
gradnt	*	*	*	*	*	*

---

note :   * = feature was derived for initial optimal feature space

@ = feature was derived for final optimal feature space

Percentage Classification Accuracy of Training and Test
Data for Analyses 17, 21, 25, 21*3, 25*5, 3225, 21*r for
Image Field (individual class covariance matrices)

Analysis:	17		21		25		21*3	
Class	Train	Test	Train	Test	Train	Test	Train	Test
1	100.0	66.6	100.0	60.0	100.0	53.3	100.0	66.6
2	98.8	83.3	96.6	70.0	100.0	86.6	100.0	83.3
3	99.2	90.0	100.0	90.0	96.5	83.3	99.7	90.0
4	100.0	73.3	100.0	76.6	100.0	66.6	100.0	80.0
5	92.8	26.6	99.5	46.6	100.0	60.0	99.5	40.0
6	100.0	43.3	100.0	20.0	100.0	13.3	100.0	16.6
7	100.0	90.0	100.0	100.0	100.0	80.0	100.0	76.6

Analysis:	25*5		3225		21*r
Class	Train	Test	Train	Test	Test
1	100.0	50.0	100.0	60.0	76.6
2	100.0	73.3	100.0	46.6	84.6
3	99.8	90.0	100.0	96.6	78.9
4	100.0	66.6	100.0	83.3	57.3
5	100.0	60.0	100.0	23.3	28.6
6	100.0	10.0	100.0	10.0	27.0
7	100.0	100.0	100.0	66.6	100.0

APPENDIX B-3

Summary of Individual Feature Subsets for Analyses 21 and 25 for Image Field

Analyses :	21		25	
Subsets :	#1	#2	#1	#2
brightness		*		*
average	*		*	
variance	*		*	
angsecaver	*			
angsecrang		*		*
corraver	*		*	
corrang	*			*
varrang		*		*
indiffrang		*		
sumenrang		*	*	
diffenrang	*			
meascoranga	*		*	
meascoraverb			*	
dvarg				*
maxprang				*
sreimrang		*	*	
contrast2		*		
glnaver	*			
gradient		*		*

note : * = feature was selected within appropriate analysis subset

APPENDIX B-4

Percentage Classification Accuracy of Training and Test Data for Analysis Subsets 21 and 25 for Image Field
(individual class covariance matrices)

Analyses :	21				25			
Subsets :	#1		#2		#1		#2	
Class	Train	Test	Train	Test	Train	Test	Train	Test
1	100.0	66.6	99.1	50.0	100.0	50.0	93.7	80.0
2	100.0	56.6	98.3	73.3	100.0	96.6	98.3	50.0
3	98.3	83.3	92.9	86.6	99.4	96.6	51.2	26.6
4	100.0	80.0	99.5	50.0	100.0	66.6	73.8	30.0
5	100.0	43.3	90.5	46.6	100.0	40.0	98.6	43.3
6	100.0	20.0	100.0	20.0	100.0	13.3	100.0	13.3
7	100.0	100.0	100.0	83.3	100.0	76.6	99.1	83.3

Percentage Classification Accuracy of Computer Selected
Test Data (ERDAS) using Final Optimal Feature Space
Classifier (FOFSC) and Computer Generated Classifier
for Analyses 17 and 21 over Image Field

Classifier Class	FOFSC 17	FOFSC 21	ERDAS 21(7)	ERDAS 21(8)
1	50.0	55.0	60.0	55.0
2	45.0	50.0	45.0	65.0
3	70.0	70.0	55.0	50.0
4	50.0	65.0	55.0	55.0
5	30.0	35.0	30.0	30.0
6	25.0	25.0	45.0	55.0
7	75.0	85.0	30.0	30.0

note : FOFSC = Final Optimal Feature Space Classifier

ERDAS(7) = Computer Generated Classifier using 7 Final
Optimal Textural Feature Images

ERDAS(8) = Computer Generated Classifier using 7 Final
Optimal Textural Images plus Textural Image
derived from feature Bright

Best Combinations of Features for Analyses 17, 21, 25,

21*3, 25*5, 3225 for Image Field

Analysis 17 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f3, f6, f13

best 4 = f2, f3, f6, f13

best 5 = f2, f3, f4, f6, f13

best 6 = f2, f3, f5, f6, f9, f13

best 7 = f2, f3, f4, f5, f6, f9, f13

best 8 = f2, f3, f4, f5, f6, f8, f9, f13

best 9 = f2, f3, f4, f5, f6, f8, f9, f13, f14

best 10 = f2, f3, f4, f5, f6, f8, f9, f11, f12, f13

best 11 = f2, f3, f4, f5, f6, f8, f9, f11, f12, f13, f14

best 12 = f2, f3, f4, f5, f6, f7, f8, f9, f11, f12, f13, f14

best 13 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14,

      f15

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f15 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = angular second moment range

    f6 = contrast average

    f7 = correlation range

    f8 = variance range

    f9 = inverse difference range

    f10 = sum entropy range

    f11 = difference entropy range

    f12 = measure of correlation A range

    f13 = measure of correlation B average

Run Length Statistics

    f14 = short run emphasis inverse moment range

Analysis 21 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f3, f6

best 3 = f3, f6, f14

best 4 = f2, f3, f6, f14

best 5 = f2, f3, f4, f6, f14

best 6 = f2, f3, f4, f6, f7, f14

best 7 = f2, f3, f4, f6, f11, f12, f14

best 8 = f2, f3, f4, f6, f7, f11, f12, f14

best 9 = f2, f3, f4, f6, f7, f11, f12, f14, f15

best 10 = f2, f3, f4, f6, f7, f9, f11, f12, f14, f15

best 11 = f2, f3, f4, f5, f6, f7, f9, f11, f12, f14, f15

best 12 = f2, f3, f4, f5, f6, f7, f9, f10, f11, f12, f14, f15

best 13 = f2, f3, f4, f5, f6, f7, f8, f9, f11, f12, f13, f14, f15

best 14 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

best 15 = f1, f2, f3, f4, f5, f6, f8, f9, f10, f11, f12, f13, f14, f15, f16

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f16 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = angular second moment range

    f6 = correlation average

    f7 = correlation range

    f8 = variance range

    f9 = inverse difference range

    f10 = sum entropy range

    f11 = difference entropy range

    f12 = measure of correlation A range

Run Length Statistics

    f13 = short run emphasis inverse moment range

    f14 = grey level non-uniformity average

Grey Level Difference Statistics

    f15 = contrast

Analysis 25 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f3, f10

best 3 = f3, f5, f10

best 4 = f3, f5, f8, f10

best 5 = f2, f3, f5, f10, f13

best 6 = f2, f3, f5, f9, f10, f13

best 7 = f2, f3, f5, f8, f9, f10, f13

best 8 = f2, f3, f5, f7, f8, f9, f10, f13

best 9 = f2, f3, f5, f7, f8, f9, f10, f12, f13

best 10 = f1, f2, f3, f5, f7, f8, f9, f10, f12, f13

best 11 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f12, f13

best 12 = f1, f2, f3, f4, f6, f7, f8, f9, f10, f12, f13, f14

best 13 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f12, f13, f14

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f14 = gradient

Co-occurrence Type Features

    f4 = angular second moment range

    f5 = correlation average

    f6 = correlation range

    f7 = variance range

    f8 = sum entropy range

    f9 = measure of correlation A range

    f10 = measure of correlation B average

    f11 = difference variance range

    f12 = maximum probability range

Run Length Statistics

    f13 = short run emphasis inverse moment range

Analysis 21*3 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f3, f6

best 3 = f3, f6, f14

best 4 = f3, f4, f6, f14

best 5 = f2, f3, f4, f6, f14

best 6 = f2, f3, f4, f6, f14, f15

best 7 = f2, f3, f4, f6, f7, f14, f15

best 8 = f2, f3, f4, f6, f7, f13, f14, f16

best 9 = f2, f3, f4, f6, f7, f12, f13, f14, f16

best 10 = f2, f3, f4, f5, f6, f7, f12, f13, f14, f15

best 11 = f2, f3, f4, f5, f6, f7, f12, f13, f14, f15, f16

best 12 = f2, f3, f4, f5, f6, f7, f11, f12, f13, f14, f15, f16

best 13 = f2, f3, f4, f5, f6, f7, f8, f11, f12, f13, f14, f15, f16

best 14 = f2, f3, f4, f5, f6, f7, f8, f10, f11, f12, f13, f14, f15, f16

best 15 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16

best 16 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17

best 17 = f1, f2, f3, f4, f5, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17, f18

best 18 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17, f18

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f18 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = angular second moment range

    f6 = correlation average

    f7 = correlation range

    f8 = variance range

    f9 = inverse difference range

    f10 = sum entropy range

    f11 = entropy range

    f12 = difference entropy range

    f13 = measure of correlation A range

    f14 = measure of correlation B average

Run Length Statistics

    f15 = short run emphasis inverse moment range

    f16 = grey level non-uniformity range

Grey Level Difference Statistics

f17 = contrast

Analysis 25*5 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f6

best 3 = f2, f3, f13

best 4 = f2, f3, f13, f15

best 5 = f2, f3, f4, f13, f15

best 6 = f2, f3, f5, f6, f7, f13

best 7 = f2, f3, f5, f6, f7, f13, f15

best 8 = f2, f3, f5, f6, f7, f9, f13, f15

best 9 = f2, f3, f5, f6, f7, f9, f11, f13, f15

best 10 = f2, f3, f4, f5, f6, f7, f11, f13, f14, f15

best 11 = f2, f3, f4, f5, f6, f7, f9, f10, f11, f13, f15

best 12 = f2, f3, f4, f5, f6, f7, f9, f10, f11, f13, f14, f15

best 13 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f13, f15

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f13, f14, f15

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16

where :

First Order Statistics :

  f1 = brightness

  f2 = average brightness

  f3 = variance of brightness

  f16 = gradient

Co-occurrence Type Features

  f4 = angular second moment range

  f5 = contrast range

  f6 = correlation average

  f7 = correlation range

  f8 = variance range

  f9 = inverse difference range

  f10 = entropy range

  f11 = difference entropy range

  f12 = measure of correlation A range

  f13 = measure of correlation B average

Run Length Statistics

  f14 = short run emphasis inverse moment range

  f15 = grey level non-uniformity range

Analysis 3225 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f5

best 3 = f2, f3, f5

best 4 = f2, f4, f5, f10

best 5 = f2, f3, f4, f15, f10

best 6 = f2, f4, f5, f9, f10, f14

best 7 = f2, f4, f5, f8, f9, f10, f14

best 8 = f2, f4, f5, f7, f8, f9, f10, f14

best 9 = f2, f4, f5, f7, f8, f9, f10, f13, f14

best 10 = f2, f4, f5, f7, f8, f9, f10, f12, f13, f14

best 11 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f12, f14

best 12 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f12, f14, f15

best 13 = f1, f2, f3, f4, f5, f6, f7, f8, f10, f12, f13, f14, f15

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f10, f11, f12, f13, f14, f15

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14,

   f15

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f15 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = correlation average

    f6 = correlation range

    f7 = variance range

    f8 = inverse difference range

    f9 = sum entropy range

    f10 = measure of correlation A range
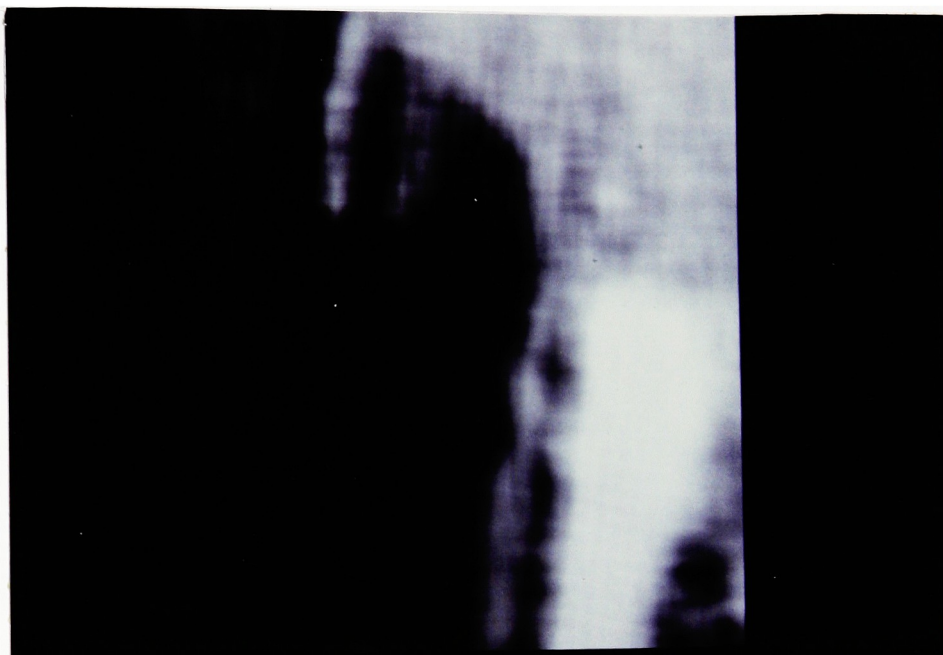
    f11 = difference variance range

    f12 = maximum probability range

Run Length Statistics

    f13 = long run emphasis inverse moment range

    f14 = short run emphasis inverse moment range

Average   Brightness   (Moyenne)

Variance   (Ecart)
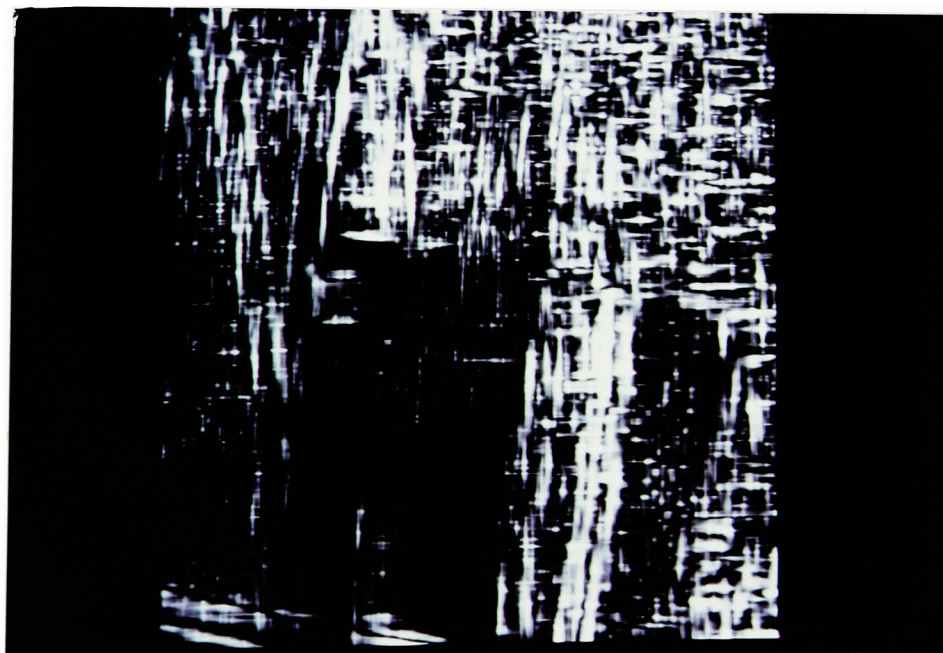
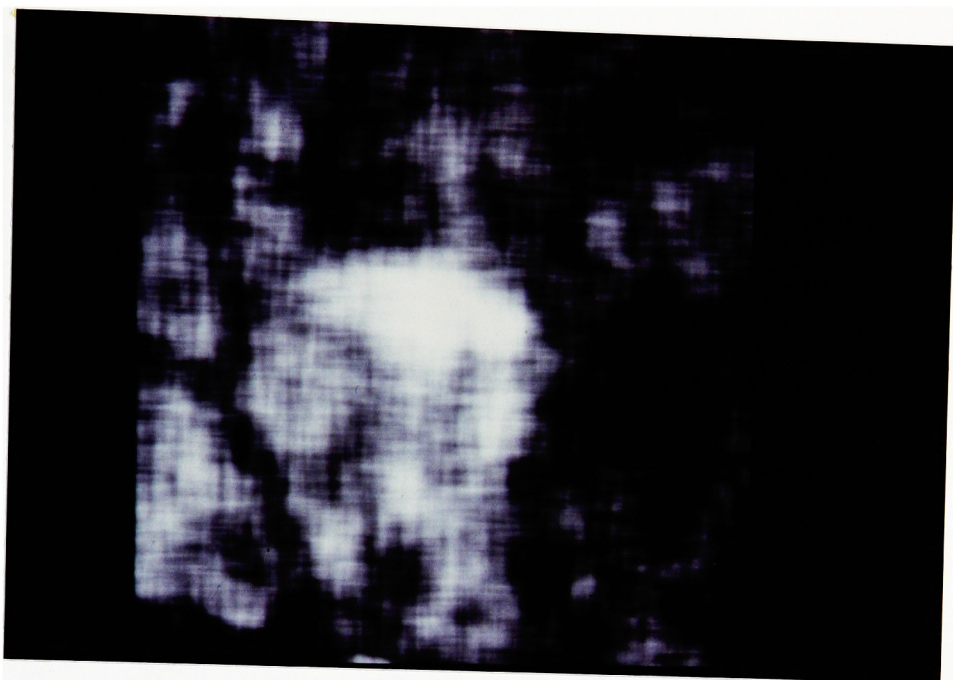**Angular Second Moment Average (Angsecaver)**

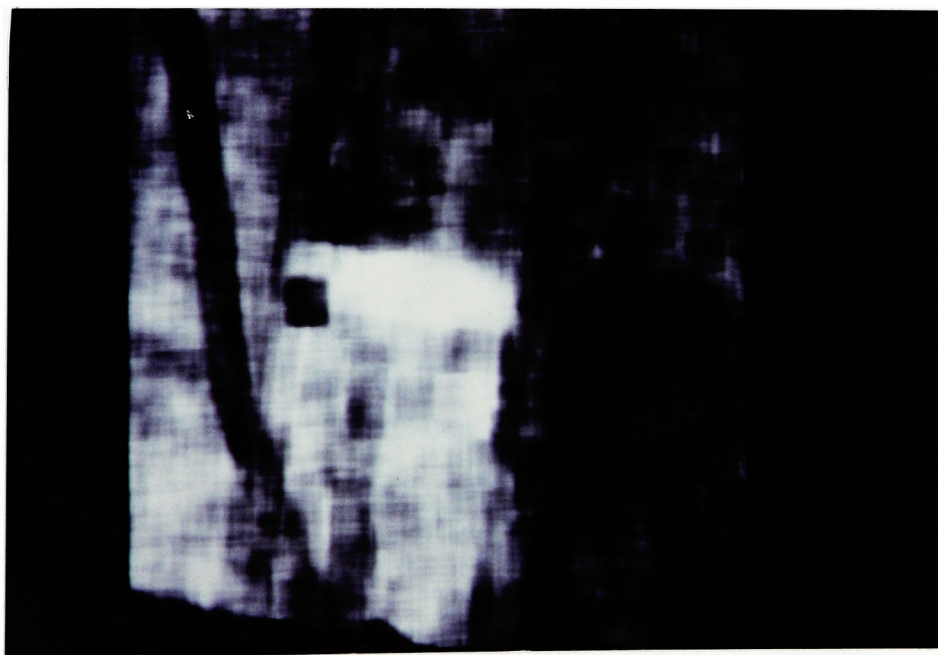**Correlation Average (Corraver)**

Correlation Range (Corrang)

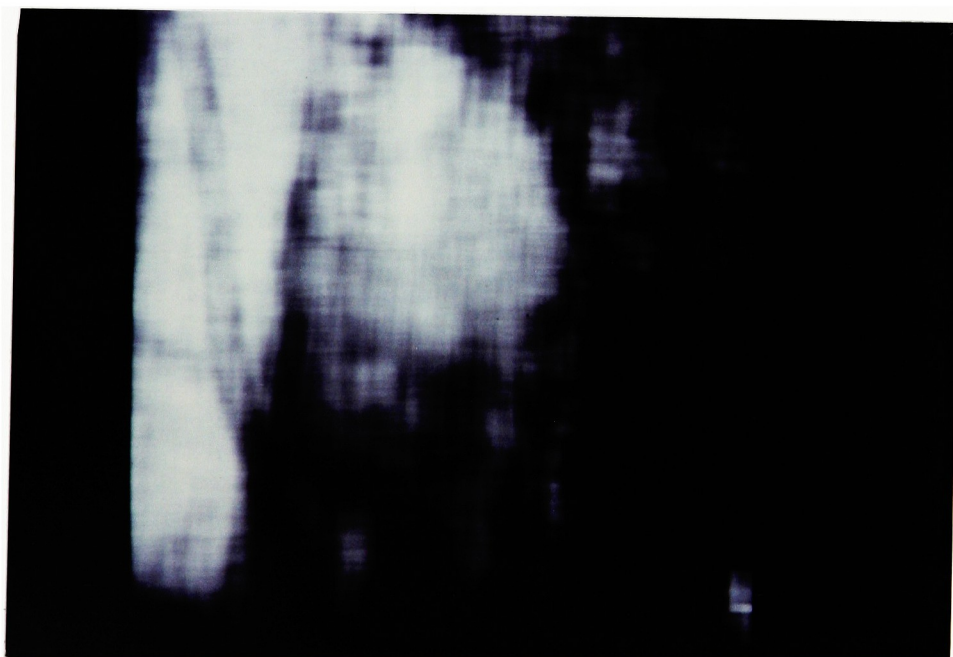**Variance Range (Varrang)**

Inverse Difference Moments Range (Indiffrang)

Measure of Correlation A Range (Meascoranga)

Maximum Probability Range (Maxprang)

Contrast (Contrast2)

Initial and Final Optimal Feature Spaces for Analyses 21, 25, 31, 3221, 3231, 31*r for Image Forest

Analysis :	21	25	31/31*5	3221	3231
Features					
brightness	*	*	*	*	*
average	*@	*@	*@	*@	*@
variance	*@	*@	*@	*	*
angsecaver	*@	*@	*@	*@	*@
contrang	*	*@		*	
corraver	*@	*@	*@		
corrang	*	*@	*@	*	*@
varrang	*@	*@	*@	*	*@
indiffrang	*@	*@	*@	*@	*@
sumenrang	*@	*@	*	*@	*@
sumvarang	*	*@		*	
entrang	*@	*@	*	*@	*
diffenrang					*@
meascoravera		*@			
meascoranga			*@		
meascoraverb	*@			*@	*@
dvarg	*@	*		*	
maxprang	*@	*@	*@	*	*@
glnrang					*@
ent2	*@	*@		*@	
contrast2			*@		*
gradnt	*	*	*	*	*

note : * = feature was derived for initial optimal feature space

@ = feature was derived for final optimal feature space

APPENDIX C-2

Percentage Classification Accuracy of Training and Test Data for Analyses 21, 25, 31, 3221 ,3231, 31*r for Image Forest (individual class covariance matrices)

Analysis:	21		25		31	
Class	Train	Test	Train	Test	Train	Test
1	100.0	76.6	100.0	73.3	100.0	70.0
2	100.0	90.0	100.0	93.3	100.0	93.3
3	100.0	83.3	100.0	53.3	100.0	46.6
4	99.9	56.6	100.0	66.6	100.0	56.6
5	100.0	90.0	100.0	96.6	100.0	96.6

Analysis:	3221		3231		31*r
Class	Train	Test	Train	Test	Test
1	100.0	80.0	100.0	76.6	66.6
2	100.0	93.3	100.0	90.0	96.6
3	100.0	76.6	100.0	30.0	26.6
4	100.0	76.6	100.0	70.0	40.0
5	100.0	90.0	100.0	96.6	100.0

APPENDIX C-3

Summary of Individual Feature Subsets for Analyses 21, 25, 31 for Image Forest

Analyses :	21		25		31	
Subsets :	#1	#2	#1	#2	#1	#2
brightness		*				*
average	*		*		*	
variance	*		*		*	
angsecaver	*		*		*	
contrang		*		*		
corraver	*		*		*	
corrang		*		*	*	
varrang		*		*		*
indiffrang		*	*			*
entrang		*		*		*
meascoravera			*			
meascoranga					*	
meascoraverb	*					
maxprang	*			*		*
sumvarang		*		*		
sumenrang	*		*			*
dvarg	*		*			
ent2		*		*		
contrast2					*	
gradient				*		*

note : * = feature was selected within appropriate analysis subset

APPENDIX C-4

Percentage Classification Accuracy of Training and Test Data for Analysis Subsets 21, 25, 31 for Image Forest
(individual class covariance matrices)

Analyses:	21				25			
Subsets:	**#1**		**#2**		**#1**		**#2**	
Class	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
1	100.0	76.6	100.0	83.3	100.0	80.0	100.0	83.3
2	100.0	90.0	100.0	93.3	100.0	83.3	100.0	93.3
3	100.0	80.0	100.0	40.0	100.0	80.0	98.4	33.3
4	100.0	56.6	97.3	50.0	100.0	56.6	99.8	66.6
5	100.0	90.0	92.7	90.0	100.0	90.0	98.4	100.0

Analysis:	31			
Subsets:	**#1**		**#2**	
Class	**Train**	**Test**	**Train**	**Test**
1	100.0	73.3	95.2	63.3
2	100.0	93.3	91.3	30.0
3	100.0	46.6	100.0	83.3
4	100.0	56.6	100.0	60.0
5	100.0	96.6	100.0	73.3

Appendix C-5

Best Combinations of Features for Analyses 21, 25, 31, 3221, 3231 for Image Forest

---

Analysis 21 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f2, f4, f6

best 4 = f2, f3, f4, f6

best 5 = f2, f3, f4, f6, f15

best 6 = f2, f3, f4, f6, f13, f15

best 7 = f2, f3, f4, f6, f13, f15, f16

best 8 = f2, f3, f4, f6, f11, f13, f14, f15

best 9 = f2, f3, f4, f6, f11, f13, f14, f15, f16

best 10 = f2, f3, f4, f6, f9, f11, f13, f14, f15, f16

best 11 = f2, f3, f4, f6, f9, f11, f12, f13, f14, f15, f16

best 12 = f2, f3, f4, f6, f8, f9, f11, f12, f13, f14, f15, f16

best 13 = f2, f3, f4, f6, f7, f8, f9, f11, f12, f13, f14, f15, f16

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f12, f13, f15, f16

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f16

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f16, f17

best 17 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17

where :

First Order Statistics :

f1 = brightness

f2 = average brightness

f3 = variance of brightness

f17 = gradient

Co-occurrence Type Features

f4 = angular second moment average

f5 = contrast range

f6 = correlation average

f7 = correlation range

f8 = variance range

f9 = inverse difference range

f10 = sum variance range

f11 = sum entropy range

f12 = entropy range

f13 = measure of correlation B average

f14 = difference variance range

f15 = maximum probability range

Grey Level Difference Statistics

f16 = entropy

Analysis 25 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f2, f4, f6

best 4 = f2, f3, f4, f6

best 5 = f2, f3, f4, f6, f9

best 6 = f2, f3, f4, f6, f9, f11

best 7 = f2, f3, f4, f6, f9, f11, f14

best 8 = f2, f3, f4, f6, f9, f11, f13, f14

best 9 = f2, f3, f4, f6, f9, f11, f12, f13, f16

best 10 = f2, f3, f4, f6, f9, f11, f12, f13, f14, f16

best 11 = f2, f3, f4, f6, f8, f9, f11, f12, f13, f15, f16

best 12 = f2, f3, f4, f5, f6, f8, f10, f11, f12, f13, f15, f16

best 13 = f2, f3, f4, f5, f6, f7, f8, f10, f11, f12, f13, f15, f16

best 14 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f16

best 15 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f16, f17

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f15, f16, f17

best 17 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16, f17

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f17 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = contrast range

    f6 = correlation average

    f7 = correlation range

    f8 = variance range

    f9 = inverse difference range

    f10 = sum variance range

    f11 = sum entropy range

    f12 = entropy range

    f13 = measure of correlation A average

    f14 = difference variance range

    f15 = maximum probability range

Grey Level Difference Statistics

    f16 = entropy

Analysis 31 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f3, f4, f5

best 4 = f2, f3, f4, f5

best 5 = f2, f3, f4, f5, f6

best 6 = f2, f3, f4, f5, f6, f13

best 7 = f2, f3, f4, f5, f6, f11, f13

best 8 = f2, f3, f4, f5, f8, f9, f10, f12

best 9 = f2, f3, f4, f5, f7, f8, f9, f10, f12

best 10 = f2, f3, f4, f5, f6, f7, f8, f11, f12, f13

best 11 = f2, f3, f4, f5, f6, f7, f8, f9, f11, f12, f13

best 12 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13

best 13 = f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

where :

First Order Statistics :

f1 = brightness

f2 = average brightness

f3 = variance of brightness

f14 = gradient

Co-occurrence Type Features

f4 = angular second moment average

f5 = correlation average

f6 = correlation range

f7 = variance range

f8 = inverse difference range

f9 = sum entropy range

f10 = entropy range

f11 = measure of correlation A range

f12 = maximum probability range

Grey Level Difference Statistics

f13 = contrast

Analysis 3221 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f2, f4, f12

best 4 = f2, f4, f12, f15

best 5 = f2, f4, f8, f12, f15

best 6 = f2, f4, f8, f11, f12, f15

best 7 = f2, f4, f8, f10, f11, f12, f15

best 8 = f2, f4, f6, f8, f10, f11, f12, f15

best 9 = f2, f4, f6, f8, f10, f11, f12, f14, f15

best 10 = f2, f4, f6, f7, f8, f10, f11, f12, f14, f15

best 11 = f2, f3, f5, f6, f7, f8, f9, f11, f12, f14, f15

best 12 = f1, f2, f3, f5, f6, f7, f8, f9, f11, f12, f14, f15

best 13 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f11, f12, f14, f15

best 14 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12 f14, f15

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f14, f15, f16

best 16 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15, f16

where :

First Order Statistics :

    f1 = brightness

    f2 = average brightness

    f3 = variance of brightness

    f16 = gradient

Co-occurrence Type Features

    f4 = angular second moment average

    f5 = contrast range

    f6 = correlation range

    f7 = variance range

    f8 = inverse difference range

    f9 = sum variance range

    f10 = sum entropy range

    f11 = entropy range

    f12 = measure of correlation B average

    f13 = difference variance range

    f14 = maximum probability range

Grey Level Difference Statistics

    f15 = entropy

Analysis 3231 Best Combinations of Features in Final Optimal Feature Space :

best 2 = f2, f4

best 3 = f2, f4, f11

best 4 = f2, f4, f11, f12

best 5 = f2, f4, f5, f11, f12

best 6 = f2, f4, f8, f10, f11, f12

best 7 = f2, f4, f8, f9, f10, f11, f12

best 8 = f2, f4, f8, f9, f10, f11, f12, f13

best 9 = f2, f4, f5, f8, f9, f10, f11, f12, f13

best 10 = f2, f4, f5, f7, f8, f9, f10, f11, f12, f13

best 11 = f2, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13

best 12 = f2, f3, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

best 13 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14

best 14 = f1, f2, f3, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

best 15 = f1, f2, f3, f4, f5, f6, f7, f8, f9, f10, f11, f12, f13, f14, f15

where :

First Order Statistics :

f1 = brightness

f2 = average brightness

f3 = variance of brightness

f15 = gradient

Co-occurrence Type Features

f4 = angular second moment average

f5 = correlation range

f6 = variance range

f7 = inverse difference range

f8 = sum entropy range

f9 = entropy range

f10 = difference entropy range

f11 = measure of correlation B average

f12 = maximum probability range

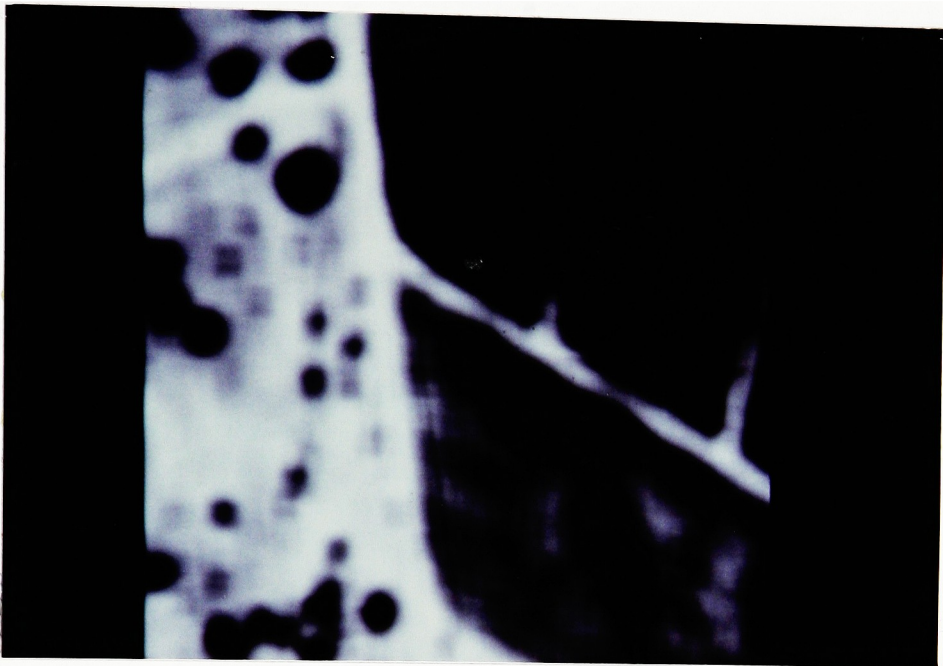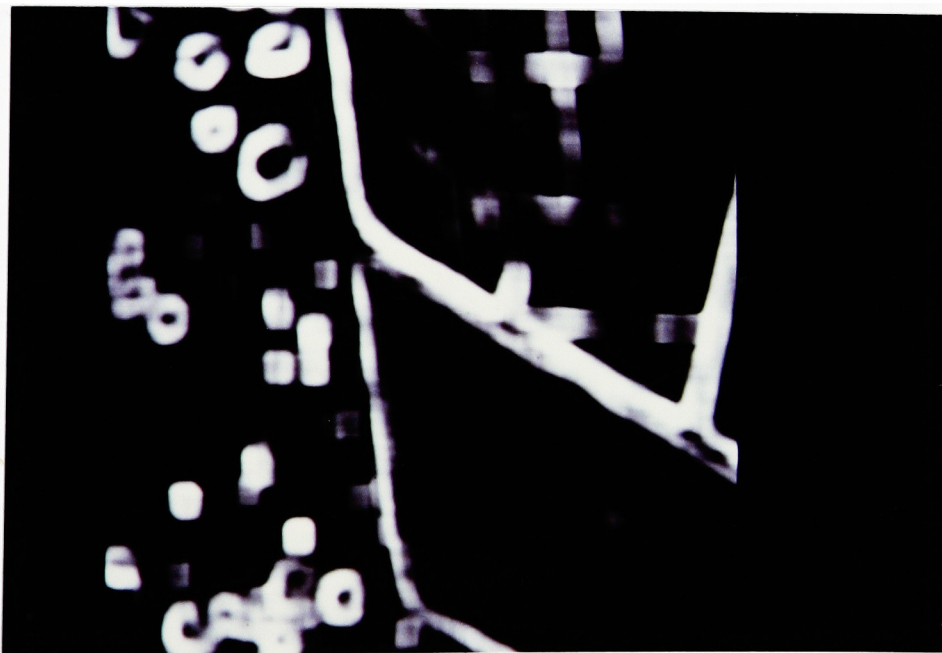## Run Length Statistics

f13 = grey level non-uniformity range

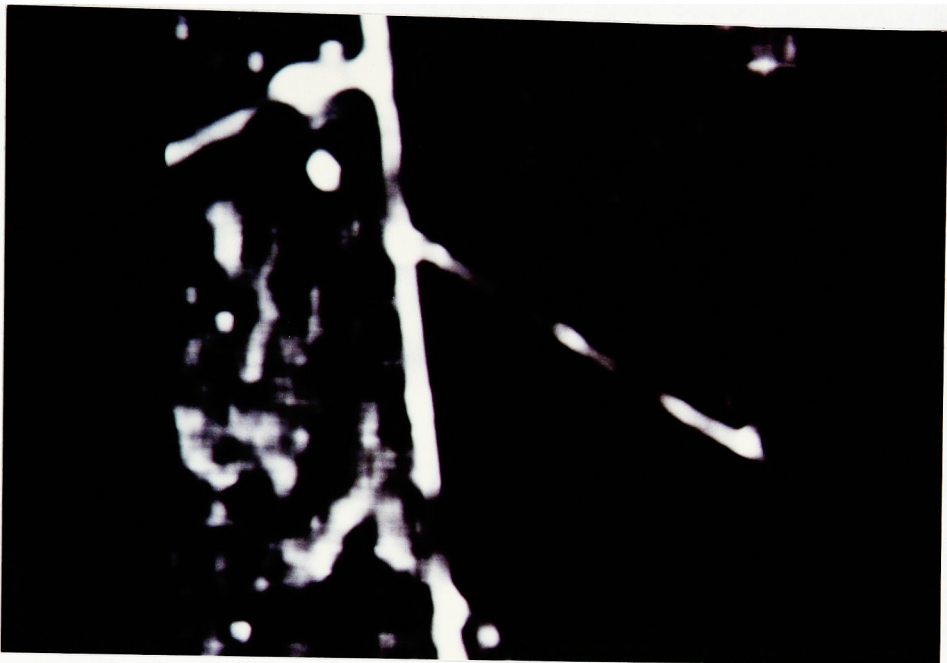## Grey Level Difference Statistics
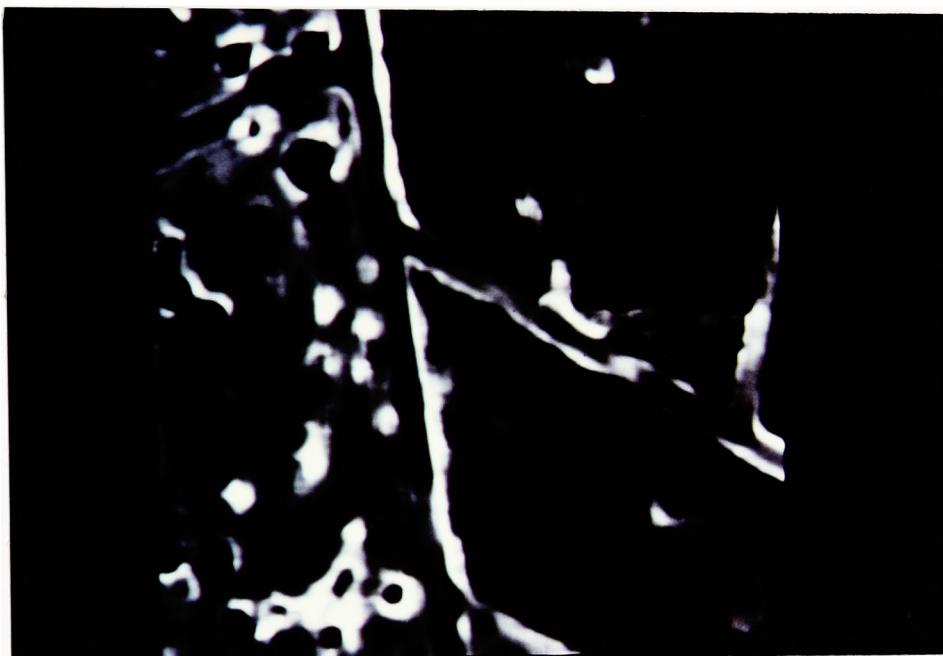
f14 = contrast

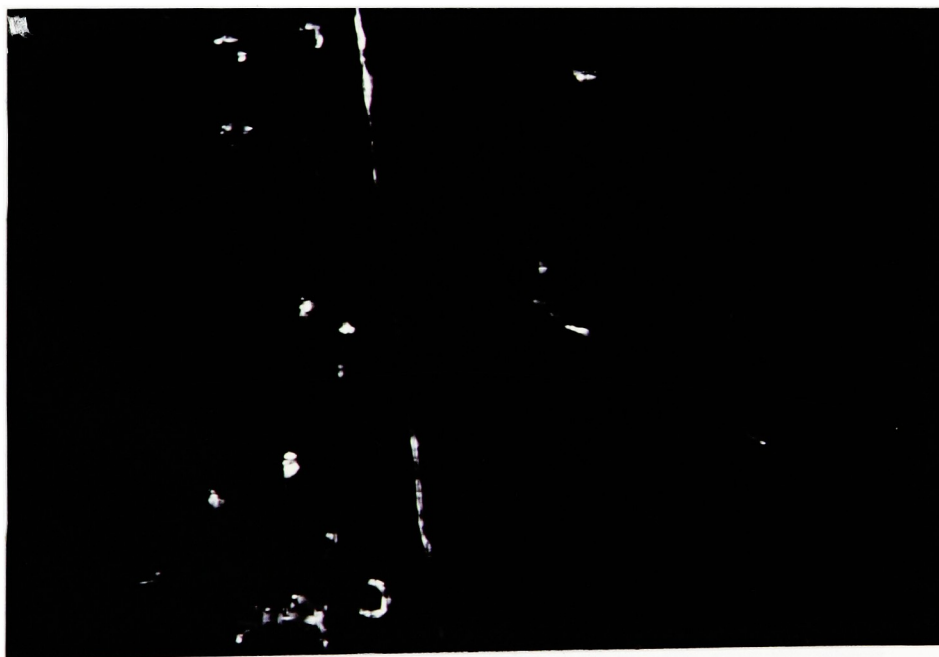Average Brightness (Moyenne)

Variance (Ecart)

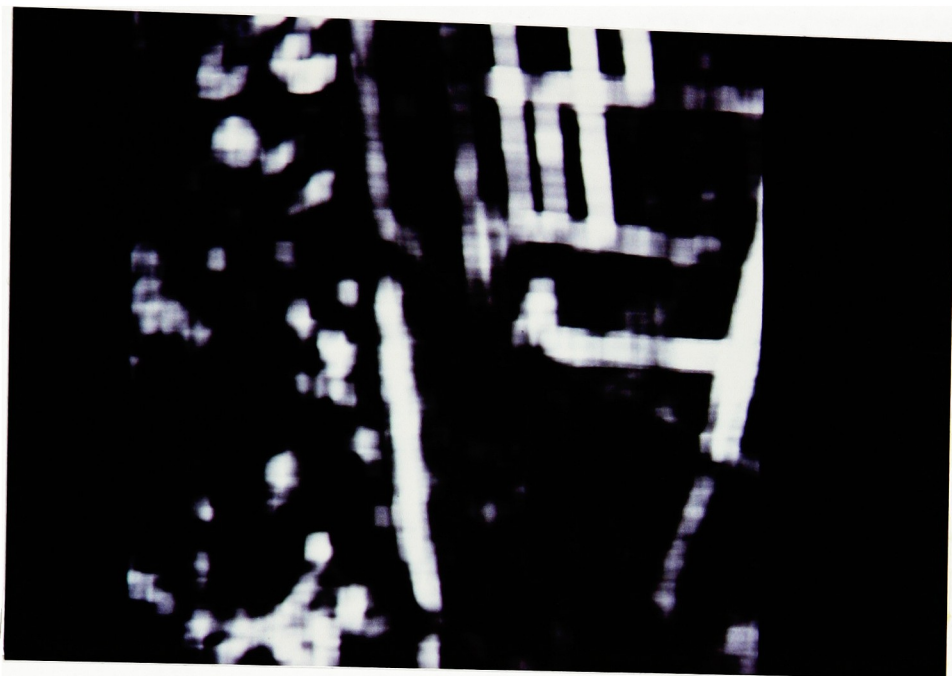Angular Second Moment Average (Angsecaver)
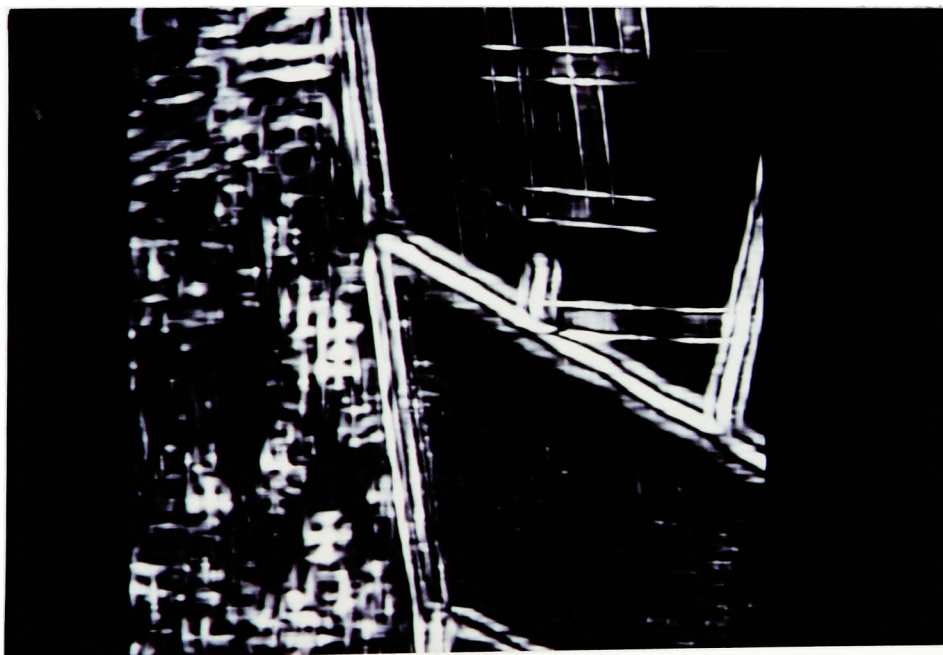

Angular Second Moment Range (Angsecrang)

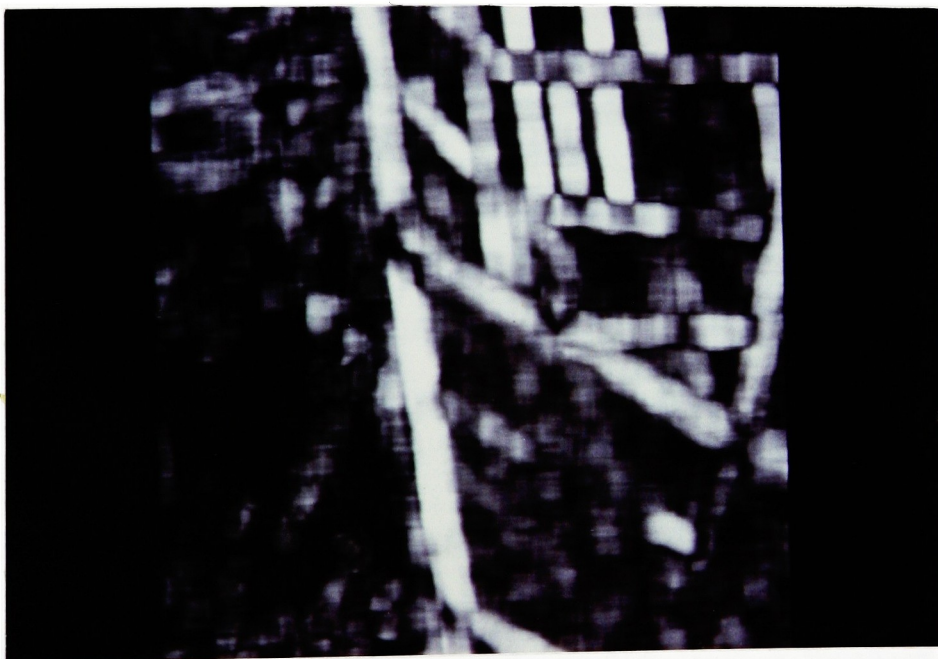Correlation Average (Corraver)

Correlation Range (Corrang)
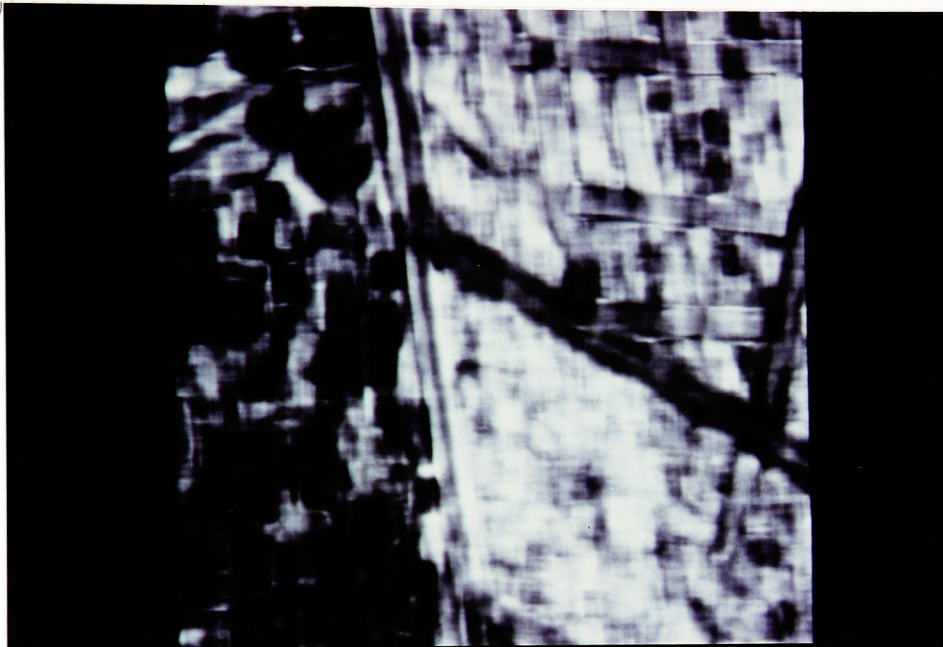
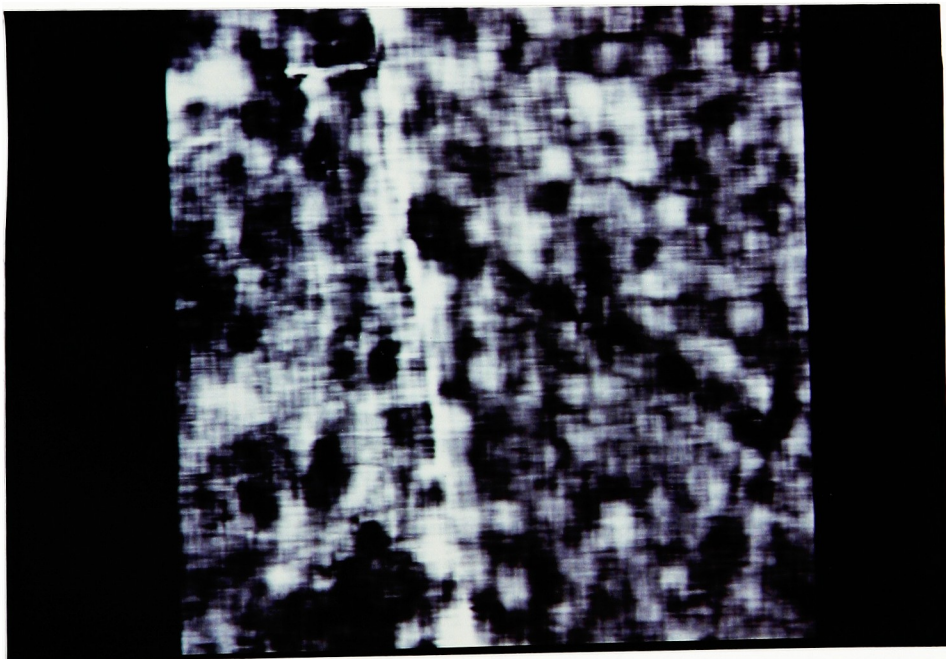Contrast (Contrast2)

Variance Range (Varrang)

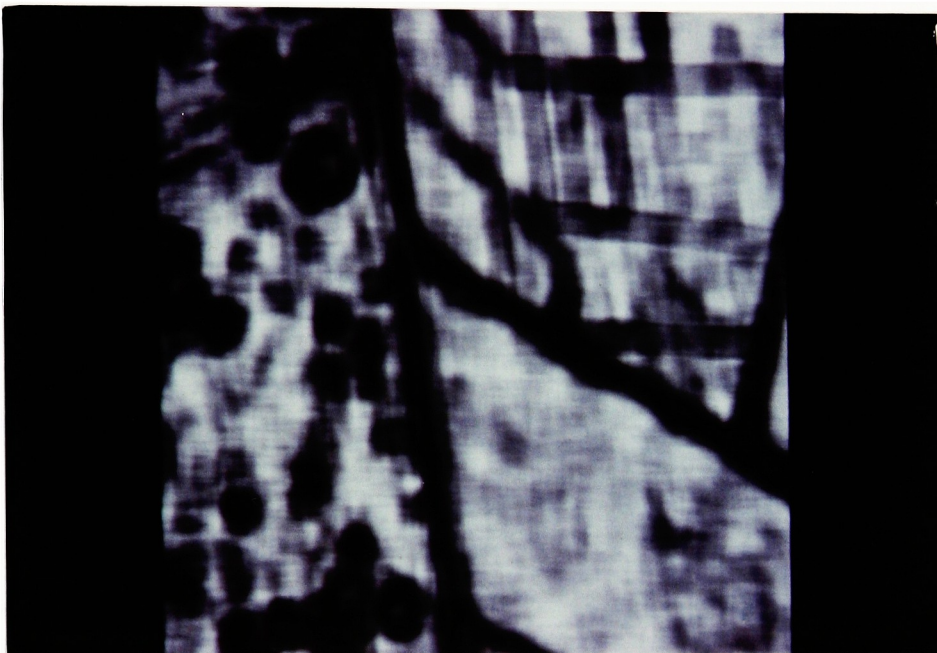**Inverse Difference Moments Range (Indiffrang)**

Difference Entropy Range (Diffenrang)

Measure of Correlation A Range (Meascoranga)

Short Run Emphasis Inverse Moment Range (Sreimrang)
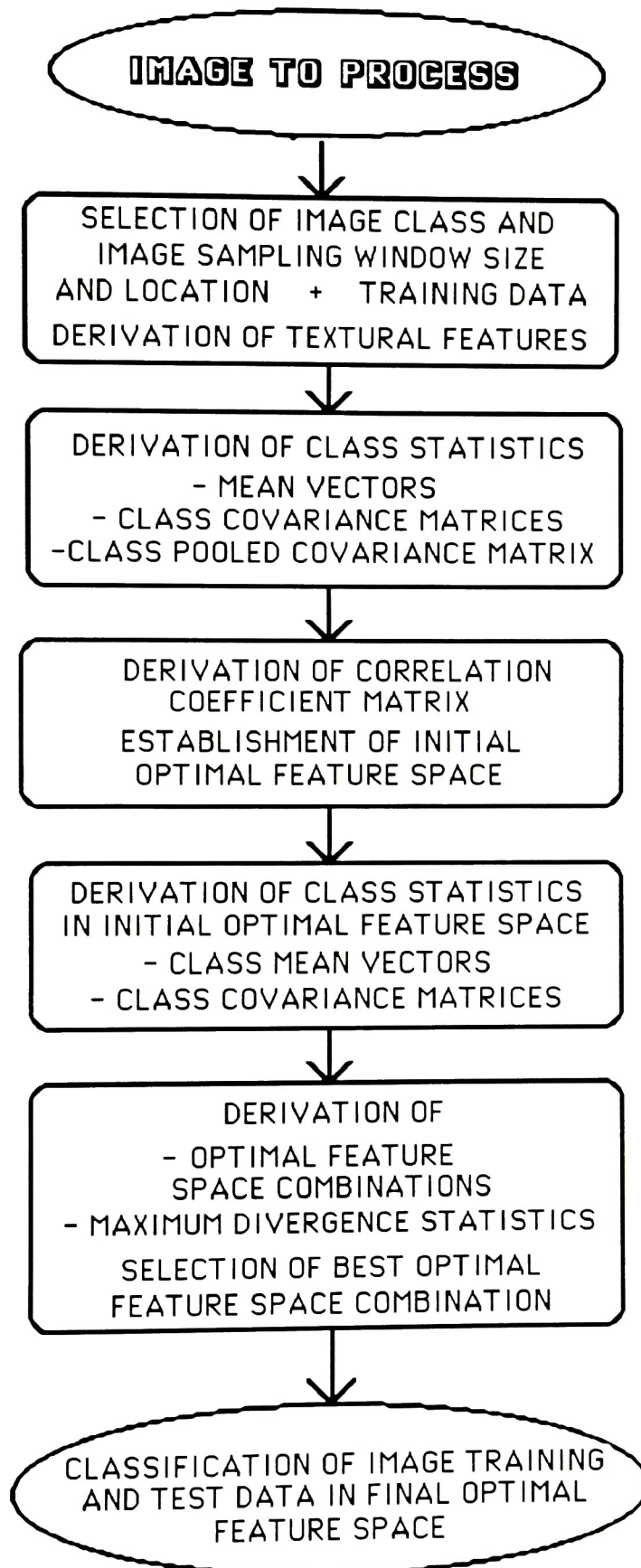
Grey Level Non-Uniformity Average (Glnaver)

Appendix D

To compute the Moore-Penrose generalized inverse, the singular value decomposition of matrix A is derived. This consists of an NxN orthogonal matrix U, a PxP orthogonal matrix V and a diagonal matrix $E = \text{diag}(var_1, \ldots, var_m)$, $m = \min(N, P)$, such that $U^tAV = [E, 0]$ if N smaller or equal to P and $U^tAV = [E, 0]^t$ if N greater or equal to P. Only the first P columns of U are computed. The rank of A is computed by counting the number of nonnegligible var_i.

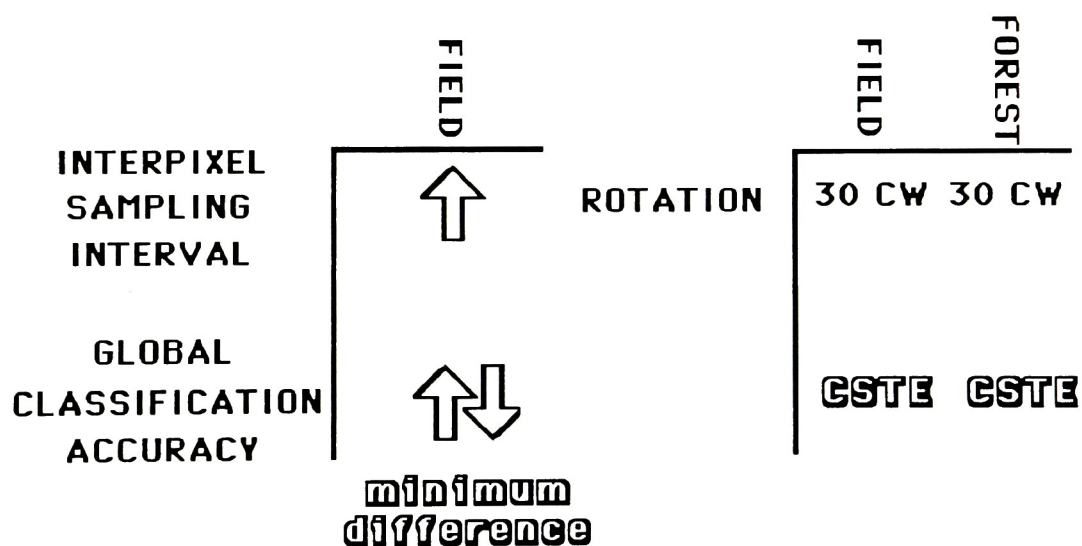The matrices U and V can be partitioned as $U = (U_1, U_2)$ and $V = (V_1, V_2)$, where both U_1 and V_1 are matrices who's dimensions are the rank of matrix A. The Moore-Penrose generalized inverse of A is

$$A_{gen\ inv} = V_1 E_1^{-1} U_1^t$$

IMAGE TO PROCESS

SELECTION OF IMAGE CLASS AND
IMAGE SAMPLING WINDOW SIZE
AND LOCATION + TRAINING DATA
DERIVATION OF TEXTURAL FEATURES

DERIVATION OF CLASS STATISTICS
- MEAN VECTORS
- CLASS COVARIANCE MATRICES
-CLASS POOLED COVARIANCE MATRIX

DERIVATION OF CORRELATION
COEFFICIENT MATRIX
ESTABLISHMENT OF INITIAL
OPTIMAL FEATURE SPACE

DERIVATION OF CLASS STATISTICS
IN INITIAL OPTIMAL FEATURE SPACE
- CLASS MEAN VECTORS
- CLASS COVARIANCE MATRICES

DERIVATION OF
- OPTIMAL FEATURE
SPACE COMBINATIONS
- MAXIMUM DIVERGENCE STATISTICS
SELECTION OF BEST OPTIMAL
FEATURE SPACE COMBINATION

CLASSIFICATION OF IMAGE TRAINING
AND TEST DATA IN FINAL OPTIMAL
FEATURE SPACE

Appendix F

	URBAN	URBAN	FIELD	FIELD	FOREST	FOREST
QUANTIZATION LEVEL	CSTE	↑	↑	CSTE	CSTE	↑
SAMPLING WINDOW SIZE	↑	CSTE	CSTE	↑	↑	CSTE
GLOBAL CLASSIFICATION ACCURACY	↓	↑↓	↓	↓	↓	↑=

minimum difference

	FIELD
INTERPIXEL SAMPLING INTERVAL	↑
GLOBAL CLASSIFICATION ACCURACY	↑↓

minimum difference

ROTATION	FIELD	FOREST
	30 CW	30 CW
	CSTE	CSTE

Note : CSTE = remained relatively constant

↑ = increase   ↑= = increase or equal

↓ = decrease   ↑↓ = increase and decrease