The improvement of digital image classification over an urban area: A probabilistic relaxation approach

Yafei Lu
University of Nebraska at Omaha

Follow this and additional works at: http://digitalcommons.unomaha.edu/studentwork

Recommended Citation
http://digitalcommons.unomaha.edu/studentwork/587

This Thesis is brought to you for free and open access by DigitalCommons@UNO. It has been accepted for inclusion in Student Work by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.
THE IMPROVEMENT OF DIGITAL IMAGE CLASSIFICATION OVER
AN URBAN AREA: A PROBABILISTIC RELAXATION APPROACH

by

Yafei Lu

A Thesis

Presented to the

Department of Geography/Geology

and the

Faculty of the Graduate College

University of Nebraska

In partial Fulfillment

of the Requirements for the Degree

Master of Arts

University of Nebraska at Omaha

December 1989
THESIS ACCEPTANCE

Accepted for the faculty of the Graduate College, University of Nebraska, in partial fulfillment of the requirements for the degree of Master of arts, University of Nebraska at Omaha.

Committee

<table>
<thead>
<tr>
<th>Name</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael P. Peterson</td>
<td>Geography/Geology</td>
</tr>
<tr>
<td>Kenneth J. Langran</td>
<td>Geography/Geology</td>
</tr>
<tr>
<td>Zhengxin Chen</td>
<td>Math/Computer Science</td>
</tr>
</tbody>
</table>

Michael P. Peterson, Chairman

Dec. 18, 1989

Date
TO ELLEN WITH LOVE
ABSTRACT

Digital image classification is a technique to extract land cover information from imagery using certain classification schemes. Additional information, either in a map, digital or verbal format, from multi-sources other than satellite imagery can be integrated into a classification scheme to improve its performance. The research in this field is essential because data interpretation has long been a weak link between the functions of two powerful systems: data acquisition by the remote sensing system and data storage, renewal and retrieval by the geographic information system - that has been more and more involved in the geographical research with the development of new concepts and technology.

Statistical models are flexible in incorporating information from multi-sources into a classification procedure to extract land cover information from satellite data. When a statistical model is applied to the urban scene, however, it is difficult to incorporate any spatial and ancillary information into the classification procedure because of their unpredictable nature in an urban scene and the difficulty in defining them into a meaningful format for a machine.

In this thesis, the probabilistic relaxation model is used with Landsat TM data of the Omaha metropolitan area. The purpose of this approach is to improve the overall classification accuracy from a Gaussian Maximum Likelihood (GML) classifier by incorporating the contextual information contained in the neighborhood of a central pixel into the classification procedure.

Prior to classification, ten land cover types determined visually from the image were combined into six classes by comparing their relative positions in a 6-dimensional classification space. After the completion of GML classification on the original image
data, the relaxation model is applied to adjust the class membership probabilities; after each adjustment, the classified map is compared with the reference map to assess the improvement of overall classification accuracy. This process is continued until the maximum accuracy is obtained. The error pattern during the iteration is analyzed and it is found that most error pixels occur at the boundaries between classes or they are located inside the dominant classes as individual parcels. Part of these errors have been corrected through the continuous adjustment of probabilities by iteration.

The rationale of all statistical models used in this study and their mathematical meanings are reviewed in this thesis. It is concluded from the actual operation of the relaxation procedure on the Landsat TM imagery of the Omaha metropolitan area that the spectral separability checking has minimized the spectral ambiguity and the relaxation model has reduced much of the identity ambiguity. The result from this combined effort is the increase of overall classification accuracy from 77.01% of the original GML classification to 88.49% after the 13th relaxation iteration. The successful operation of the procedure has shown that it has a potential to be used in a computer-assisted land use monitoring system for the purpose of resource management.
**CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Problems</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Literature Review</td>
<td>9</td>
</tr>
<tr>
<td>1.4 Purpose of the Thesis Research</td>
<td>19</td>
</tr>
<tr>
<td>1.5 Summary</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER II METHODOLOGY AND PROCEDURE</td>
<td>22</td>
</tr>
<tr>
<td>2.1 Data Preparation</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Training Area Selection</td>
<td>24</td>
</tr>
<tr>
<td>2.3 Spectral Separability Checking</td>
<td>28</td>
</tr>
<tr>
<td>2.4 Iterative Refinement of Classification</td>
<td>31</td>
</tr>
<tr>
<td>2.5 Summary</td>
<td>35</td>
</tr>
<tr>
<td>CHAPTER III STATISTICAL MODELS</td>
<td>38</td>
</tr>
<tr>
<td>3.1 Multivariate Analysis in Remote Sensing - General</td>
<td>38</td>
</tr>
<tr>
<td>3.2 Statistical Characterization of Spectral Intensities - Mean Vector and Covariance Matrix</td>
<td>40</td>
</tr>
<tr>
<td>3.3 Statistical Characterization of Class Membership - Probability</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Density Function</td>
<td>43</td>
</tr>
<tr>
<td>3.4 Determination of Class Membership - Maximum Likelihood Decision Rule</td>
<td>45</td>
</tr>
<tr>
<td>3.5 Statistical Separability - Divergence and J-M Distance</td>
<td>52</td>
</tr>
<tr>
<td>3.6 Statistical Characterization of the Contextual Information</td>
<td>59</td>
</tr>
<tr>
<td>3.7 Iterative Refinement of Classification - Probabilistic Relaxation Procedure</td>
<td>66</td>
</tr>
<tr>
<td>3.8 Accuracy Assessment - Error Matrix</td>
<td>69</td>
</tr>
<tr>
<td>3.9 Summary</td>
<td>74</td>
</tr>
<tr>
<td>CHAPTER IV TRAINING DATA MANIPULATION</td>
<td>77</td>
</tr>
<tr>
<td>4.1 On the Selection of Training Data</td>
<td>77</td>
</tr>
<tr>
<td>4.2 Matching Spectral Classes with Informational Classes</td>
<td>80</td>
</tr>
<tr>
<td>4.3 Summary</td>
<td>86</td>
</tr>
<tr>
<td>CHAPTER V ANALYSIS</td>
<td>87</td>
</tr>
<tr>
<td>5.1 Probabilistic Relaxation</td>
<td>87</td>
</tr>
<tr>
<td>5.2 Error Patterns</td>
<td>92</td>
</tr>
<tr>
<td>5.3 Summary</td>
<td>100</td>
</tr>
<tr>
<td>CHAPTER VI CONCLUSIONS</td>
<td>102</td>
</tr>
<tr>
<td>6.1 On the Classification Scheme</td>
<td>102</td>
</tr>
<tr>
<td>6.2 Some Suggestions on the Further Research</td>
<td>105</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>108</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>137</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1-1 Basic Components of A Remote Sensing Operation System ................. 3
Figure 1-2 The Variations Arising from a Single Informational Class ................... 6
Figure 1-3 Basic Problems of Digital Image Classification ................................. 7
Figure 1-4 Examples of Information Integration for Forest and Agriculture Scenes . 13
Figure 2-1 Study Area ....................................................................................... 23
Figure 2-2 Separability in the Sets Consisting of Different Measurement Parameters .......................................................... 28
Figure 2-3 Separability in Terms of Training Statistics in the Same Measurement Set ........................................................................... 29
Figure 2-4 The Basic Steps of Spectral Separability Checking ............................. 30
Figure 2-5 Operational Process of Gaussian Maximum Likelihood Classification . 32
Figure 2-6 Concept of Subjective Compatibility and Statistical Compatibility .... 33
Figure 2-7 The Steps of Updating Class Membership Probabilities Using Relaxation Procedure .......................................................................... 34
Figure 3-1 N-Band Training Data for Class i, Where $x_{jk}$ is the Gray Level on Band j of the kth pixel for class i ................................................................. 41
Figure 3-2 Two Dimensional Decision Regions and Boundaries ......................... 46
Figure 3-3 Maximum Likelihood Decision Rule .................................................. 47
Figure 3-4 Divergence and J-M Distance as Functions of Normalized Distance ...... 58
Figure 3-5 The Behavior of the R operator to be Used in Updating Probabilities .... 60
Figure 3-6 The Conversion of Probabilities to Compatibility Coefficients ............ 64
Figure 3-7 Specific Neighborhood for the Pixels Located at the Corners and Edges of the Probability File ................................................................. 65
Figure 3-8  Non-Site-Specific Accuracy Assessment ........................................... 70
Figure 3-9  Site-Specific Accuracy Assessment ..................................................... 70
Figure 4-1  Histograms of Gray Level Distribution from WATER Training Data Set . . . 79
Figure 4-1  Histograms of Gray Level Distribution from INDTR Training Data Set . . . 80
Figure 5-1  Percentage of Correct Classification After Thirteen Iterations ............. 91
Figure 5-2  WATER Coverage from GML and the Last Iteration ............................. 93
Figure 5-3  The Reduction of Error Rate by Iterative Adjustment of Probabilities . . 95
LIST OF TABLES

Table 2-1  The Initial Definition of the Land Cover Categories ...........................................27
Table 2-2  Index of the ERDAS Programs Related to This Study ........................................36
Table 3-1  Example of an Error Matrix ...................................................................................71
Table 3-2  Errors of Omission and Errors of Commission .....................................................73
Table 3-3  Notational Index .....................................................................................................75
Table 4-1  The Standard Deviation Extracted from the 6-band Training Data .....................78
Table 4-2  Training Pixel Number (TPN) for Different Land Cover Categories ....................78
Table 4-3  J-M Distance Between Different Land Cover Categories .....................................81
Table 4-4  Divergence Between Different Land Cover Categories ........................................82
Table 4-5  J-M Distance from Different Land Cover Categories After Adjusting ..................84
Table 4-6  Divergence from Different Land Cover Categories After Adjusting .....................85
Table 5-1  Error Matrix from the GML Classification .............................................................88
Table 5-2  Compatibility Coefficients from a 3 by 3 Neighborhood .....................................90
Table 5-3  Error Matrix from the Last Iteration .....................................................................92
ACKNOWLEDGEMENTS

I would like to give my sincere thanks to those who agreed to serve on my thesis committee. As chairman of the thesis committee, Dr. Michael P. Peterson has given me great encouragement in facing the challenge posed by the thesis research and was always willing to spend time with me in discussing all possible topics that eventually led to the completion of the thesis research. His advise, patience and expertise in the areas associated with my thesis research is greatly appreciated. The contribution by Dr. Kenneth J. Langran must be recognized for his experience and expertise in the application of remote sensing to geography, which is greatly influential to me in finishing this research. The suggestions from Dr. Zhengxin Chen are also very helpful in organizing the programs used in the thesis research.

Additional thanks are given to Mr. Roger M. Hubbard, manager of the Remote Sensing Application Laboratory, for his help in transferring data between different computer systems that has facilitated the hard job I would otherwise have to face.

Finally, I would like to thank the Graduate Studies and Research and the UNO Alumni Association. Without their initial support and financial help, it would be very difficult for me to come to this country and have an excellent education in this university.
CHAPTER I  INTRODUCTION

1.1 Introduction

Geography is the science that deals with the earth and life, especially the human environment. Traditionally, a map is a major information source for geographers in their research and routine work. In some sense, geography is a map-oriented science because any kind of analysis accomplished in the geographical context can be generalized in the form of map or graphics based on the classifications following certain analytical procedures. As a result of geographical research, this kind of generalization created by a classification procedure represents the acquisition of knowledge and the practical use of this knowledge in making a more comprehensive understanding of geographical phenomena and processes summarized from the raw data (Mather 1979). This is especially true in the physical domain of geography.

Remote sensing is one of the more technical approaches used in geography for the location, classification and estimation of features in the environment (Curran 1987b). During the last four decades, the development of remote sensing technology from visual interpretation of aerial photographs to the computer-aided analysis of satellite imagery has posed a great challenge to those applied geographers who are interested in the application of this technique to our discipline in meeting human needs (Curran 1987a). Although there exist so many definitions of the term "Remote Sensing", either in a broad sense (Holz 1973, Estes and Senger 1974, Lintz and Simonett 1976, Richason 1978, Barrett and Curtis 1982), or in a narrower sense (Sabins 1978, Slater 1980, Townshend 1981, Simonett 1983, Lo 1986, Lillesand and Kiefer 1987), and the discussion of its current position in geography (Everett and Simonett 1976, Munton and
Goudie 1984, Fussell, Rundquist and Harrington 1986, Curran 1985), the real value of remote sensing as a technique applied to geographical research is unquestionable (Bauer 1976), as realized by scientists and highlighted by numerous research works on the automated analysis done in the geo-science related disciplines beginning from 1970's (Fu 1976).

Upon the launch of the first Earth Resources Observation Satellite and the first acquisition of imagery in digital format, researchers focused on its practical use, but few could foresee the enormous stimulus it would give to the development of improved methods of information extraction using digital techniques (Dahlberg and Jensen 1986). Some of the possible advantages of digital techniques have been recognized by researchers in this field. First, it can avoid human subjectivity introduced by manual interpretation of raw data because once the the digital classification procedure is determined, the phenomena presented on the imagery is interpreted based on the universal principle defined by a particular scheme; while in manual interpretation, the different interpreters may give different results according to their opinions and understandings to the phenomena under investigation. Second, it can release the human resources by performing routine analysis procedures under human instructions. The successful application of these advantages to the environmental problems, however, depends on the successful establishment of interactions between human and machine.

Remote sensing, while a powerful component of data acquisition that provides an efficient way of observing the earth surface, is hindered by the difficulty of processing the voluminous amounts of data. This processing usually involves conventional interpretation techniques that are both time and labor intensive. With the increasing concern on the monitoring of our changing environment, it has posed a great challenge to geographers to develop more efficient and accurate data analysis procedures to facilitate decision making process for resource managers, planners and administrators. For this
purpose, tremendous efforts have been made to explore the possible application of image processing techniques to automated image analysis and employ this potential to its maximum extent by using digital classification techniques to overcome difficulty in the conventional data analysis procedure. With the development of geographical information systems from concept to operation, the geographers will have more options to utilize the information stored in this powerful data storage and retrieval system, which makes it more urgent for geographers to improve our analytical method that has long been a weak link between, and may hinder the efficient functioning of, two powerful tools, remote sensing and geographic information systems, as graphically illustrated by Figure 1-1.

Like any theory or technology in its early development, digital classification technique is far from perfect (Foody 1987). Among various numerical techniques, statistical models have been able to produce better results when applied in image classification because of their capability in imitating the process of human analysis. The classification is especially impressive when the statistical models are applied to the scenes with higher homogeneity like those of large area agriculture and forest, but unfortunately seldom tested in the heterogeneous area, typical of the urban scene.

Figure 1-1. Basic components of a remote sensing operation system.
After several years of practice, the remote sensing specialists began to realize the limitation of the statistical models and the importance of incorporating information other than spectral information into the classification process, either in a pre-classification or post-classification mode to improve the accuracy. Among these efforts, the probabilistic relaxation technique has received much attention because of its capability and flexibility to incorporate statistical or other kind of contextual information into the classification to improve a GML classifier. The main goal of this study is to try to incorporate contextual information contained in the neighborhood of a central pixel into a GML to improve its performance over an urban area using a probabilistic relaxation model. The accuracy of classification is checked repeatedly after each iteration using the ground truth data created by actual image interpretation. Hopefully, this study can further increase the practical usage of remote sensing technology, a goal that many applied scientists in geo-science related disciplines have been pursuing.

1.2 Problems

Digital image classification is the process of assigning pixels to corresponding classes based on the spectral reflectance of these pixels recorded over different bands. It plays a very important part in the automated image analysis (Jensen 1983). Though there are many alternatives to establish different kinds of classification schemes, as well as the actual numbers of categories to be discriminated by digital techniques, the determination of them is not an arbitrary one at all (Hixson et al. 1980), which has not been fully recognized by some workers in this field. Historically speaking, many statistical models currently used in image data classification are modifications and
refinements of those originally developed by statisticians in the 1930's and used in social and biological sciences. They were implemented on digital computers when they became less expensive in the 1960's to test hypotheses based on multivariate measurements of the observed phenomena (Anderson 1958, Mather 1976). The measurements of reflectance made remotely by sensors over n different bands is just the analogue to the observations over n variables as the input to a conventional multivariate statistical model to initiate the multivariate classification (Landgrebe 1976). Here, two important concepts have to be mentioned: the spectral classes and the informational classes. The informational classes are:

categories of interest to the users of data, for example, the different kinds of geological units, different kinds of forest, or the different kinds of land use that convey information to planners, managers, administrators, and scientists who use information derived from remotely sensed data. These classes form the information that we wish to derive from the data - they are objects of our analysis (Campbell 1987, pp. 296-297).

while the spectral classes are:

groups of pixels which are uniform with respect to the brightness in their several spectral channels. The analyst can observe spectral classes within remotely sensed data; if it is possible to define links between the spectral classes on the image and the informational classes that are of primary interest, then the image forms a valuable source of information (Campbell 1987, pp. 297).

Thus, spectral classes may represent completely different kinds of informational classes, such as granite in a geological unit and the roof of a commercial building that they have the same spectral behavior, which is one of the main sources of classification error. Another example is that a region of the informational class "residential" is still
"residential", even though it may display variations in housing development, composition, and density, all of which may or may not form distinct spectral subclasses, depending on if they are spectrally separable from each other and thus are able to match their corresponding informational classes (Figure 1-2).

![Diagram of Informational Class and Spectral Subclasses](image)

**Figure 1-2.** The variations arising from a single informational class.

Unfortunately, the match between spectral and informational classes is not easy to make. On one hand, it is important to define spectral classes into their corresponding informational classes, and on the other hand, it is also important to establish informational classes to which the spectral classes can be possibly and meaningfully defined. In other words, only those informational classes that are both meaningful to the user and spectrally separable from each other in an multi-dimensional classification space should be established. Too many informational classes may make the classification impractical and too few informational classes may lose information that is of interest to
the user. This point can be illustrated in Figure 1-3. In Figure 1-3a, the homogeneous patterns of classes are distinct in terms of training statistics and thus the spectral classes can be matched to their corresponding informational classes without ambiguity. With the heterogeneous area in Figure 1-3b, however, it is almost impossible to solve the problem of ambiguity represented by the overlay area 2 using only spectral information, because two different informational classes are in fact represented by the same spectral pattern.

![Figure 1-3](image)

a. Homogeneous training data.  
b. Heterogeneous training data.

Figure 1-3. Basic problems of digital image classification.

So, before any supervised classification is initialized, the first question to be asked is what kind of information one is going to use as the input to a particular classifier and what output from the classification represents. If the spectral information is the major source of information, we have to make sure that the spectral classes corresponding to the output informational classes from a classification scheme present reasonable
spectral separability in an n-dimensional spectral space. Only by starting from this point can we proceed to solve the problem represented by the overlap area 1 in the Figure 1-3b.

Most classification schemes based on statistical models proceed on a pixel by pixel basis and thus have the same deficiency of exploiting only spectral information and being unable to incorporate other information contained in the relationship between each pixel and its neighboring pixels - a common practice in human interpretation of an image. As the effort of increasing classification accuracy, spatial information has been gaining more and more attention in the past few years. The concept of spatial information, however, is only a broadly defined term, ranging from image texture derived from the pattern of reflectance occurrence over the entire image to the spatial relationships between a specific geographical location and the object of primary interest. These kinds of patterns are sometimes hard to be defined in a raster format compatible with digital imagery (Schowengerdt 1983, Ekstrom 1984).

Ancillary information that has long been used in manual interpretation represents the spatial information in a narrower and more limited sense, including the information from higher resolution remote sensing products, maps, reports, and even personal experience (Fox, Brockhaus and Tosta 1985). Some of the ancillary information are in raster formats, like the digital elevation models stored in a geographic information system and the imagery with other special spectral characteristics (Liu, Teng, and Xiao 1987), and are easily incorporated into a classification. Some others are not, like the distance between features of geographical interest and the convenience of access to an irrigation facility. These kinds of information are often in a very generalized form and is difficult to be explicitly defined (Harris 1981, Burrough 1986).

The contextual information refers more specifically to the information contained in the neighborhood of a pixel, representing different information defined by different
classification procedures (Swain, Siegal and Smith 1980), though there might be some other interpretations of contextual information when it is used in different contexts by different users (Wharton 1982). All of these kinds of spatial information can be incorporated into classification in a pre-, during, or post-classification mode depending on the nature of the tasks at hand (Hutchinson 1982, Gurney and Townshend 1983).

As a spectral classifier, the GML functions fairly well in summarizing the spectral characteristics of a particular spectral group during the training stage, but the reliability often declines at the decision making stage. The probable reason for this declining is that the GML has a very strict rule of assigning a pixel to a specific informational class according to maximum likelihood decision rule during the decision making stage, without considering the contextual information at all. In other words, the probability of its neighboring pixels' being assigned to certain classes can not play a role in the pixel assignment. This rule is employed unconditionally to every pixel throughout the whole classification, even when ambiguity arises between different classes - the ambiguity represented by the almost equal likelihood in terms of the probability density function, typically of the situation shown by the overlap area 2 in Figure 1-3b. This is the main problem to be approached by this study.

1.3 Literature Review

One of the discriminant techniques suggested for use in analyzing satellite data is the Gaussian maximum likelihood classifier. With this classifier, one assumes that spectral intensities of pixels come from the multispectral measurements within the Instantaneous Field Of View (IFOV) and form normal distributions. Data from pixels contained in training sets are used to produce spectral means and associated co-variance
matrices for each type of land use category. Each pixel is then allocated to its corresponding informational class based on the spectral characteristics summarized in the training stage (Kershaw 1987). This technique has a strong theoretical basis in statistical theory (Nillson 1965) and it has been successful in agriculture, forest inventory, and other remote sensing applications (MacDonald et al. 1972, Bizzell et al. 1975). Its degree of success, however, is not always consistent through the practical remote sensing applications, especially when attempted in detailed classification of an urban area using low resolution MSS data, or when used in identifying detailed land covers in areas with medium to highly diversified spectral intensities using a high resolution TM data (Khorram, Brockhans and Cheshire 1987, Haack 1987).

After some practical applications of the GML to the real remotely sensed data during 1970's, the attention was then switched to the computational improvements and algorithm evaluation of probability and the maximum likelihood decision rule to increase its efficiency and accuracy. Such efforts include the development of hybrid classifiers which use parallelepiped algorithms first and then turn to maximum likelihood computation to resolve ambiguity (Goodenough and Shlien 1974), the look-up table scheme to reduce the repeated calculations (Shlien and Smith 1975), the incorporation of prior probabilities as weighting functions into final classification to increase accuracy (Strahler 1980), and the different strategies of decision rules (Tom and Miller 1984). It is surprising to notice, however, that the link between spectral and informational classes seems to have been missing as a research topic for quite a long time (Campbell 1987), which could probably be attributed to the poor accuracy that resulted at times from automated analysis. In fact, one of the important steps involved in classification procedure like the GML supervised classifier like GML is to determine first the informational classes with the data available at hand and then try to match the spectral classes with the corresponding informational classes before the classification is
actually done; otherwise, it is only an unsupervised classifier to group pixels together based on reflectance with less degree of human control, losing the advantages of a supervised classifier. While comparing probability of error is not often feasible, we can reasonably predict higher error rates when pixels are allocated to those informational classes such that their corresponding spectral classes are in fact not separable in an n-dimensional classification space.

There are two basic measurements of statistical separability. Divergence is used to measure statistical separability in pattern recognition (Marill and Green 1963) and Jeffries-Matusita (J-M) distance can measure statistical separability between pairs of classes (Swain and Davis 1978). The basic assumption of separability checking using training data is that if training data is not spectrally separable, the ground cover corresponding to this training set can not be considered as a separate informational class. The principle of this analytical procedure is quite easy to understand: while making the maximum use of a classifier, one should not try to do what a classifier is not able to do.

Different efforts have been made on different stages of classification for the purpose of accuracy improvement (Nelson 1985). Traditionally, the first step utilizing the information contained in a digital image is called pre-processing, which is a technique to generalize the image data first before starting a classification procedure to reduce the variability introduced by the sensor system (Mather 1987). An example of pre-processing is to replace a band in a particular band combination by its mean-filtered and median-filtered counterparts. This procedure may reduce the scene noise and classification errors that occurred among the more heterogeneous class patterns introduced by high spatial resolution remote sensing products, provided that the scale of heterogeneity is smaller than that of the filter (Atkinson, Cushnie and Townshend 1985). Another example is the radiometric correction in mountainous area. Woodham
and Gray (1987) demonstrated in an experiment the relationship between the reflectance and terrain and calibrated on the reflectance of the shadowed area using a model more general and powerful than those used in current operational systems to compensate the effect of topography over spectral reflectance before proceeding to classification, which may remove the shadow effect produced under the specific sun angle and the slope orientation.

Post-classification manipulation is a technique to make generalization after the classification is done in order to reduce variability introduced during the procedure. The method of data manipulation depends on what degree of generalization is going to be achieved. An example of post-processing is presented by Townsend (1986). A smoothing operator is designed to move over the entire classified image with a salt-and-pepper appearance to remove "noise" while the behavior of the operator is confined in order not to lose useful information.

Another kind of approach to improve accuracy is to incorporate such ancillary information as topographical, geological, soil, climatological, and ecological, into the classification systems (Hutchinson 1982). These kinds of data are highly correlated with the formation of land covers in some specific areas, having either positive or negative influences on the overall spectral reflectance of these land covers (Miller and Shasby 1982, Cibula and Nyquist 1987). For example, in a high mountain area, different species of forest are not spectrally separable, but they are differentiated simply by thresholding the elevation above that some of species cannot live ecologically (Strahler, Logan and Bryant 1978). These kinds of information integration can work only in such areas that they really contribute to the differences in terms of spectral intensity under different topographical, climatological and ecological conditions (Figure 1-4). Note that in Figure 1-4, the ancillary information outlined by the double frames does have influence on the development of the land covers growing in the scenes under
consideration. Its usefulness, however, is subject to vary, depending on human's
definition on how much influence it may have on a particular environment.

Figure 1-4. Examples of information integration for forest and agriculture scenes.

When the ancillary information is applied to the classification of an urban scene,
however, the difficulty faced by information integration is that an urban area is usually
developed under such moderate ecological conditions that the environment may generally
have nothing to do with or contribute little influence on the spectral reflectance of the
existing land covers. It is a reasonable assumption because no urban area is expected to
have developed on a rugged mountainous area, without transportation facilities to its hinterland, and under tough ecological conditions (Yeates and Garner 1982).

An urban area is a complicated system with a highly diversified pattern of development. Some other considerations should be included in the classification procedure when trying to apply digital techniques to the areas with such a feature. Jensen conducted a study of utilizing textural features to address problems of image classification of urban scenes and concluded that the textural features derived from spectral intensities tend to change quickly over a short distance or within a small area and the weighting function has to be chosen carefully for the purpose of accuracy improvement (Jensen 1979); this feature can be incorporated with the original data in classification to detect the local land use change of an urban system (Jensen and Toll 1982). In his study of the image of Omaha metropolitan area, Li found that different local spatial features are present among the new and old residential areas and the new terminology is introduced to describe this spatial phenomena. This measurability of spatial features may have the potential to differentiate the new residential areas from the old ones in digital classification, though the technology of incorporating the spatial features into the classification was not attempted (Li 1987).

Given the fact that the spatial information is hard to be defined over the entire imagery and ancillary information can not play an important rule, we need a different model to solve the identity ambiguity problem. This model should behave such that at the both sides of the boundaries between two major classes, the distinctiveness, or in terms of the GML, the probabilities associated with the membership of two major classes, will stay the same and distinct while the probabilities associated with the membership of pixels at the boundary and within a dominant class are "adjusted" by taking into account the probabilities of their neighboring pixels (Peleg 1981, Harris 1985). This is another way that the information contained in the neighborhood of a pixel is exploited.
The general use of contextual information has been investigated by a number of researchers and the advances achieved in 1970's have been surveyed by Fu and Mui (1981) and Landgrebe (1981), among which many of the discussions are focused on the performance of maximum likelihood classification procedure. One of the disadvantages of this procedure is that once the classification is initiated, it is hard to incorporate additional information into decision making process to improve its performance. As an effort to solve this problem, the probabilistic relaxation model was first introduced by Rosenfeld, Hummel and Zucker (1976) into scene labeling, following the work by Waltz (1975), to deal with the identity ambiguity problem by incorporating contextual information into the classification system. In their paper, the concept of compatibility coefficient was proposed to measure statistical compatibility between different classes using probabilities of neighboring pixels. Roughly speaking, relaxation is an operation that "relaxes" the maximum likelihood decision rule. A relaxation operator should be able to behave such that the large part of the original information is retained until the very last step when the final decision of the pixel membership is made (Kittler and Foglein 1984). The major advantage of this technique over pre- or post-classification processing is that the information on how close a pixel is to other classes is measured. Comparing pre- and post-processings, the former often introduces too much unwanted smoothing and thus losing some important information through generalization (Davis and Rosenfeld 1978, Gilmour 1987) while the later uses only the class information that the pixel has been assigned to, like that done by Townsend (1986).

One of the most well known contextual classifiers is the Extraction and Classification of Homogeneous Object (ECHO) discussed by Ketting and Landgrebe (1976). The ECHO first divides the scene into homogeneous segments and then classifies these segments as units by an extended version of GML algorithm based on statistics extracted from trainable classes and neighboring pixels. Anuta et al. (1984) tested
ECHO on both MSS and TM imagery to evaluate the improvement of classification accuracy by incorporating contextual information and reported an overall increase of accuracy over an agriculture scene from 67.4% on MSS four bands using GML per pixel classifier to 97.9% on TM seven bands using ECHO.

The relaxation procedure is done in an iterated way to "adjust" repeatedly the class membership probabilities of a pixel located at the specific position, usually the central position, of a local operator. Eklundh, Yamamoto and Rosenfeld (1980) applied the probabilistic relaxation model to adjust the initial probabilities iteratively over the entire image, reporting a decrease of error rate by 1.5% after 19 iterations and having observed that most of the error occurred at the corners and boundaries between classes. Richards, Landgrebe and Swain (1981a) proposed a theoretical scheme to improve the relaxation procedure by assuming that the initial probability has major influence on the relaxation throughout the whole process; during the iteration, this influence is always exerted in a proper manner on the probabilities under modification. The same authors (1982) also conducted a study to develop the concept and technique of incorporating ancillary information in the form of a priori probability \( \Omega_i(\beta) \) into the kth iteration to modify the standard probabilistic relaxation procedure, in which the elevation data in a rugged mountain area was used as ancillary information to supervise iteration and an increase of forest classification accuracy from 68% to 81% was reported.

The accuracy of probabilistic relaxation is dependent on the number of iterations. Richards, Landgrebe and Swain (1981b) reported an observed behavior of the relaxation model that the error rate tends to drop quickly after the first few iterations and reaches the minimum after certain number of iterations and then may increase with further iterations, depending on the data sets used. This topic is further discussed by Zenzo et al (1987a) and the technique of controlling number of iteration is described and tested using both artificial imagery and real remotely sensed data.
Another interesting study was approached by Swain, Vardeman and Tilton (1981) in which the numbers of iteration and the dimension of central pixel neighborhood are related to each other in finding the context contribution and in improving the classification accuracy over simulated and actual remote sensing data. For the simulated data, the best classification on the first iteration was obtained by the second nearest neighbor; on the third iteration, the fourth nearest neighbor was deemed the best; and finally at the seventh iteration, the eight nearest neighbor produced the best result. For the real remote sensing data consisting of 2,500 pixels, the first iteration based on the northern neighbor (one nearest neighbor) resulted in increased accuracy from 83.1% to 84.2%; the second iteration based on the neighbors to the north and east gave 1% more accuracy; and the eight nearest neighbor produced the most excellent result up to 93.8% accuracy. Here, though the determination of direction in selecting neighboring pixels is somewhat arbitrary, it still can give an idea of incorporating different context under different situations, depending on the prior understanding to the scene under consideration (Tilton, Vardeman and Swain 1982). As a practical approach, Zenzo et al (1987b) utilized four iterations of probabilistic relaxation followed by four iterations of fuzzy relaxation using eight nearest neighbors to achieve an overall improved accuracy from 72.2% to 82.0% for MSS data of an agriculture scene and concluded that most of the benefits to be obtained by relaxation are concentrated on the first several iterations and the accuracy is either converged to a fixed point or degraded with further iterations. For the fairly recent surveys on the development of probabilistic relaxation technique over the past decade, see Nagy (1984), Kittler and Pairman (1985).

The accuracy of any interpretation of an image, either done by machine or manually, is the primary concern of the user of that result (Ginevan 1979, Story and Congalton 1986). Yet the questions concerning accuracy are quite difficult to address and many of the methods for accuracy estimation are themselves subject to error. The errors are not
distributed randomly over the image, they often occur at the edge of parcels or the interiors of parcels, representing their specific spatial relationships with these parcels. The most difficult question to address, however, still remains the classic one: the accuracy of the reference map (Steven 1987). Unfortunately, the plain truth is that there is no reference map that is a hundred percent accurate. The mapping process is a very complicated one and the chances of errors being introduced into the system are anywhere throughout the operation, beginning from data acquisition to the final stage of data compilation and production. To be more practical, Hay (1979) suggested that different sampling techniques be based on different error patterns in approximating the error estimation. He identified five major questions to be asked in the accuracy assessment:

1. What proportions of all decisions are correct?
2. What proportion of the allocation to a category is correct?
3. What proportion of the true category is correctly allocated?
4. Is a category over-estimated or under-estimated?
5. Are the errors randomly distributed?

While there are many sampling strategies available to address the accuracy assessment problems that have occurred in digital image classification, the universal principle is quite simple: using the best image and ground truth data available at hand to create a reference map or data as accurately as possible and then establish an efficient and convincing procedure to compare the classification map with reference map or data to assess the actual value of different classification schemes (Hord and Brooner 1976). In fact, this method of accuracy assessment is the comparison of machine-made classification with the map that is achieved by using physical and human resources
available at hand. This map is supposed to be the best base map for the purpose of comparison. When such a reference map or data is available, this is the best way to avoid human subjectivity introduced by selecting any sampling design strategy for the purpose of error estimation.

1.4 Purpose of the Thesis Research

Through the literature review, it can be observed that the probabilistic relaxation technique has played an important role in the improvement of supervised classification. Its value rests on its capability and flexibility of using either contextual information contained in neighboring pixels or ancillary information that is available in the form of compatibility coefficients, and its ability in the iterative refinement of classification to achieve the maximum benefit of a GML. The rationale of this technique is valid in the terms of statistical theory and its value when applied to remotely sensed data of agriculture and forest environments has been revealed by many researchers from different disciplines, and yet its possible potential application to urban scene has not been investigated in a systematic way. Two major problems approached by this study from the view of the practical use of Landsat imagery to urban land use characterization are:

1. The spectral ambiguity: the assignment of pixels to actual land use categories is not the simple matching of pixels to their corresponding land use categories contained in a land use map according to their spectral intensities; instead, it is first of all a spectral characterization process using statistics extracted from the training data before proceeding to classify them. This thesis will attempt to solve the problem of
spectral ambiguity reduction by first using a spectral separability checking strategy before proceeding to any classification procedure.

2. The identity ambiguity problem: the rate of mis-assignment of pixels by a GML classifier can be reduced by incorporating the statistical characterization of the contextual information contained in the neighborhood of a central pixel. This thesis will try to solve this problem by clarifying the identity ambiguity using the probabilistic relaxation procedure. The accuracy after each refinement of classification is assessed by means of error matrix.

1.5 Summary

Remote sensing technology has provided scientists a more efficient way to observe the earth that allows the classification of the observed entities into categories, which eventually serves as the basis of decision making by resource managers. The development of this technique over the past several years has changed many aspects of scientific methods used in the geographical research and the enormous amount of information brought by this technique has posed a serious challenge to geographers on how to efficiently use the available information to increase our understanding of the world. From the technical point of view, the pace of remote sensing data processing can hardly catch up with the speed of data acquisition and the information contained in the raw data is far from being efficiently utilized to its potential. Tremendous efforts have been made to extract information by automated techniques to make up this gap. Partial success has been achieved for forest and agricultural scenes, but not so much for urban scene, partly because of the highly diversified nature of the urban scene in terms of
spectral intensities and partly the difficulty in exploiting ancillary information related with the urban environment. Though the automated image analysis technique is far from perfect, like any theory and technique in their early development, one thing is certain that it is important to understand the phenomena presented on imagery using as much information as possible before the accurate and efficient analysis can be achieved. This is an important step in the context of geographical research.

Any kind of information extraction technique is subject to error, or the mis-assignment of pixels by a classification procedure. The errors may come from two major sources: the attempt to classify pixels based on its spectral reflectance over several bands to some categories while these categories can not be separated in an n-dimensional classification space because of the limited spectral resolutions. In other words, the informational classes are established based on wrong assumption. Another major source of the errors is the statistical model itself; the models which can use only spectral information of the individual pixel without referring at least the spectral characteristics contained in the context of that pixel tend to produce poorer result than that produced by those which have the flexibility in incorporating the information contained in the neighborhood of that particular pixel other than using only the spectral information of that individual pixel itself.

The spectral ambiguity between different informational classes can be measured quantitatively by spectral separability checking before the classification is actually done, reducing greatly the risk of errors from the first source. The identity ambiguity of a pixel in classification can be reduced by the probabilistic relaxation technique, which may make the risk of errors from the second source as low as possible. The performances of post-classification and the iterative relaxation can be compared with the reference map and the problems of error rate and patterns will be addressed in an appropriate way.
CHAPTER II METHODOLOGY AND PROCEDURE

2.1 Data Preparation

The quad 3 of the un-georeferenced Landsat TM image path 028 and row 031 (scene ID Y5047816360X0) taken in June 22, 1985 covering the Omaha metropolitan area is the main data used in this study. The TM imagery provides advantages with its higher spectral and spatial resolution over the earlier MSS data. Poor spatial resolution typically represented by MSS data creates too much generalization and a higher rate of classification errors; while the higher spatial resolution data like an aerial photograph introduces too much spatial variation (Markham and Barker 1985). The moderate spatial resolution of 30 meters represented by TM imagery is suitable to conduct a study with much consideration of spectral variation than that of spatial variation. The existing land use map compiled by the Department of Omaha City Planning in 1986 and the high resolution near-infrared aerial photo taken in 1980 are used to compile the reference map, combined by field study when necessary. All data resided on the SUN workstation are transferred to the VAX of the University of Nebraska at Omaha Campus Computing through tape media for data manipulation. The entire study area consists of a 512 by 512 portion of the quad 3 and is shown in Figure 2-1. The portion within the dotted frame is the area to be classified by the procedure described in this chapter.

The main reason in choosing this area is that this area consists of major urban land cover types that have much spectral and spatial diversity than those areas consisting of homogeneous land cover categories. This characteristic is the major source of the classification error when using any classification procedure. The entire area is almost covered by high resolution color infrared aerial photo that is precise enough to be used
as ground truth reference. The major land cover categories of the city to be classified can be visually depicted from both a TM image and an aerial photo for the purpose of comparison and reference map compilation. The inspection of the TM image and aerial photo shows that there are no major changes of the land cover over the time interval between that when the image and aerial photo were taken.

A portion of the image within the dotted frame of Figure 2-1 is manually segmented

Figure 2-1. Study area.
by outlining those land cover categories, to be determined following the procedure
described in the section 2.3, with polygons on the display device using ERDAS* program
DIGSCR. The output .DIG file is then converted into raster format using ERDAS program
VECRAS. The output .GIS file from this program is the reference map to be compared
with the classified image for the assessment of accuracy improvement after each
refinement of classification. Though this is not the "100% correct" reference map in
terms of perfectness, it is the best base map obtained using the available data and device
at hand. The advantage of this method is that there is no mis-registration between
ground truth and the original data set, its disadvantage is that it is subject to the loss of
accuracy owing to the difficulty in controlling the movement of the cursor, especially
when the parcel is substantially small. A subimage of 256 by 256 pixels, or an area of
4.77 miles by 4.77 miles, that has all the land cover categories to be classified is taken
as test set by the relaxation procedure, considering the work load involved in manual
segmentation and the disk space available in the host computer. Some basic statistics
such as the total pixel numbers of each land cover category and the percentage of each
class to the entire area to be classified is prepared from the subimage for the purpose of
the error matrix compilation.

2.2 Training Area Selection

The accurate numerical representation of land cover category depends on the careful
selection of training data sets. This is an essential step for a supervised classification.

* Note: for the ERDAS programs used in this study, see Table 2-2 Index of the ERDAS
Programs Related to This Study of Section 2.5 in page 36 for index.
The classification based on the bad training data may deteriorate the result that otherwise should be better achieved based on the good training data. To prevent this situation from happening, some rules should be followed when selecting training data sets:

1. A training area should be both spectrally and spatially homogeneous; if this situation is not easy to obtain in a specific scene, select those training areas that are as homogeneous as possible by trying different areas of the same land cover category. The homogeneity can be visually depicted on the imagery or tested using such basic statistics as mean and standard deviation.

2. The standard deviation derived from the training data should be as small as possible, suggesting the lower spectral variability of this training set. Theoretically speaking, the standard deviation from an entirely homogeneous training area is zero.

3. The training areas should be visually separable, either spectrally or spatially. This rule guarantees that basic separability exists among different land cover categories, while the possible use of spatial context is only in the situation that it increases the spectral separability.

4. The gray levels in a training data set should be normally distributed, which is the basic requirement of the statistical models used in this study. Though the moderate violation of this rule may not be of detriment to the following analysis, the severe violation may disable the entire analytical procedure because there is no theoretical basis in the context of non-normal distributions when considering a GML classification scheme.

5. The adequate training pixels must be available to estimate the mean vector and covariance matrix for each class. If n bands are to be used for classification, the theoretical minimum number of training pixels that constitutes a training data set
should be n+1. In practice, this number should be 10n, or even 100n to provide an accurate estimate of the statistical parameters.

Attention should be paid to the rule 3. It states the fact that when choosing a training area to extract training statistics, it does not matter if the training area is only spatially separable from other categories because we are going to "check" if it is also spectrally separable.

The basic difference between the land use map and TM imagery rests on the fact that the former puts emphasis on both ownership of land and the actual function of a particular piece of land while the latter represents the actual land cover type. The distinction is essential for the establishment of land classification system. For example, a lake with its surrounding grasses in a public park is classified into public land by the definition of a land use map, while in fact it is classified as different land cover category, for example water, by a classification scheme used to classify a Landsat image. So, the naming of land use categories of the training areas are not based on the formal land use map because the resolution of the TM imagery and study procedure are different from those used to compile the formal land use map; the definition of the classification system for the land cover categories is also different. Instead, they are defined based on the interpretation from imagery and aerial photo and the rules outlined above. The formal definitions of the initial land cover categories are listed in Table 2-1.

A quick inspection of the definitions and descriptions listed in Table 2-1 reveals that the GOLFC, GRASS and VTREE are associated with vegetation coverage; COMWE, COMCR and INDTR are related to the built-up areas without or with little vegetation cover in between, a characteristic of commercial land use. The original purpose of selecting these land cover categories as training areas is to try to establish the land classification system for the approach while the statistical separability among these land
<table>
<thead>
<tr>
<th>Land Cover Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WATER</td>
<td>Continuous coverage of water body over the entire training area.</td>
</tr>
<tr>
<td>GOLFC</td>
<td>Golf course with much homogeneous coverage of fine grass that can be visually depicted from the color infrared aerial photo but not so easily from a monoband image displayed on the black and white monitor.</td>
</tr>
<tr>
<td>GRASS</td>
<td>Large area covered by homogeneous grass.</td>
</tr>
<tr>
<td>VTREE</td>
<td>Area covered mostly by trees.</td>
</tr>
<tr>
<td>BAREG</td>
<td>Bare ground without any surface cover except soil, this is the cleared area subject to further development.</td>
</tr>
<tr>
<td>COMWE</td>
<td>Commercial land use category with a more compacted pattern. The Westroads Mall is selected as training area of this type of land use category where the buildings are more closely grouped and surrounded by parking lots.</td>
</tr>
<tr>
<td>COMCR</td>
<td>Commercial land use category with a scattered pattern. The Crossroads Mall is chosen to represent this sort of land use category where the buildings are a little bit far apart from each other.</td>
</tr>
<tr>
<td>INDTR</td>
<td>Industrial area where the buildings of different patterns and sizes are grouped together to form the industrial tract.</td>
</tr>
<tr>
<td>RDOLD</td>
<td>Residential area where the houses are more densely distributed and some big trees are scattered over the area. Usually speaking, this is the characteristic of the old residential area.</td>
</tr>
<tr>
<td>RDNEW</td>
<td>Residential area where the houses are not so densely grouped together like that of RDOLD and there are few big trees within the area. Usually speaking, this is the characteristic of the new developed residential area.</td>
</tr>
</tbody>
</table>

The initial definitions of the land cover categories is checked using the technique described in the next section to make sure that the classification scheme can reach such detail without losing too much accuracy.
2.3 Spectral Separability Checking

The establishment of land classification system does not mean that we can classify them accurately using a classification scheme that puts more emphasis on the spectral domain and the contextual information summarized from the spectral information. Without solving the spectral ambiguity problem like that of the overlap area 2 presented in Figure 1-3b, we can not proceed to solve the identity ambiguity problem with an acceptable confidence.

Figure 2-2. Separability in the sets consisting of different measurement parameters.

The solution of spectral ambiguity can be approached by checking the spectral separabilities between pairs of training samples represented by training statistics. The statistical separability checking serves two purposes. The first purpose is to select the band combination that does the best job of separating the pairs of classes which are hardest to separate. In Figure 2-2, we can see that the class 1 and class 2 are better...
differentiated in a 2-dimensional space consisting of band $X_c$ and $X_d$ than that consisting of band $X_a$ and $X_b$. Another purpose of checking is to match the spectral classes into their corresponding informational classes. In other words, in an n-dimensional space, we have to determine whether or not to combine two spectral classes into one informational class. If they are too "near" in terms of statistical separability like that of Figure 2-3a, we combine them into one class if we have reason to believe that they belong to the same informational class; or keep them as separate informational class if they are "far" enough, like the situation in Figure 2-3b. The strategy adopted in this approach is the second one.

![Figure 2-3](image)

a. Class 1 and 2 are too "near" in the set $(X_a, X_b)$ and then is combined.

b. Class 1 and 2 are "far" enough in the set $(X_a, X_b)$ and stay separate.

Figure 2-3. Separability in terms of training statistics in the same measurement set.

After the first group of land cover categories identified on the image is established by referring land use map and aerial photo, the corresponding statistics of the land cover categories are extracted. The ERDAS program SIGSUB provides a very flexible capability to manipulate training statistics extraction. By running SIGSUB, the training statistics on different band combinations over exactly the same training areas are taken
for the purpose of comparison. Output mean value and covariance matrices from SIGSUB are then sent to a computer program as the input to determine the spectral separability between different land cover categories by the procedure illustrated in Figure 2-4.

Figure 2-4. The basic steps of spectral separability checking.
Several band combinations are then tried for the purpose of selecting the best separation in this approach. Roughly speaking, the band combinations are visual to near infrared (band 2, 3, and 4), near-infrared to middle-infrared (band 4, 5, and 7), and all seven bands except band 6 (thermal) because of its poor spatial resolution.

2.4 Iterative Refinement of the Classification

As a conventional maximum likelihood classifier, it checks alternatively all of the probability density values for a pixel under current classification and then assigns the pixel to the class having the maximum probability density value according to a very strict decision rule. The essential deficiency of this approach is illustrated in the output portion of the Figure 2-5: it does not consider any possibility implied by the situation of very close and relatively high probability values. When the probabilities for different classes are so close like that of 0.36 for class 2 and 0.37 for class 3, what we are facing in fact is the identity ambiguity problem shown by the overlap area 1 of Figure 1-3b. Note that in this situation, the class membership is quite sensitive; if the probabilities changed slightly to 0.37 for class 2 and 0.36 for class 3, the class membership of the pixel under consideration is switched to class 2. Under such a sensitive situation, we need to look into the probabilities of the neighboring pixels to determine which class should be favored for the class membership assignment.

The use of contextual information consists of two parts. The first part is to summarize the class compatibility. In the manual interpretation of an image, we use subjective compatibility to make generalization. Here, the term "subjective compatibility" refers to the human opinion on whether classes are compatible or not. In our example of Figure 2-6, the incompatible nature is apparent because we know from
the common sense that there should not be a cornfield in a residential area. The most probably we do in visual classification is to put this particular pixel into residential category. The difficulty in machine classification is, however, that it is not easy to define the human subjectivity into a format that the classifier can understand and thus is able to incorporate it into classification. The statistical compatibility, on the other hand, refers the information summarized from the probabilities of neighboring pixels around the central pixel. The classification scheme has the ability itself to look actively for the clues contained in the neighboring pixels regarding the class membership by following the procedure defined by a particular statistical model. The compatibility often mentioned in the approach, however, refers to the statistical compatibility because

![Diagram](image_url)

Figure 2-5. Operational process of a Gaussian maximum likelihood classification.
it is the machine understandable information. The second part of using contextual information is the iterative refinement of classification using the statistical compatibility in the form of compatibility coefficients, the means to use the clues regarding the class membership summarized in the first part of the procedure.

![Concept of subjective compatibility and statistical compatibility](image)

**Figure 2-6.** Concept of subjective compatibility and statistical compatibility.

To make the whole idea possible, the original probabilities of each pixel's being the different land cover categories are calculated and contained in a file. The compatibility coefficients are created from this file as the input to the program to update probabilities iteratively. The compatibility coefficient matrix consists of an m by m by 9 matrix, where m denotes the actual number of land cover categories to be classified and 9 refers to the eight nearest neighbors plus the central pixel itself.

The probabilities created by the (l-1)th iteration are updated in the lth iteration, where l=1, 2, ..., s-1, s, and then the probability file is converted into actual land cover categories according to maximum likelihood decision rule. Though the decision rule is at this point still strict, the whole procedure has been "relaxed" because the decision is not made until all of contextual information contained in the neighborhood of a pixel is considered.
Figure 2-7. The steps of updating class membership probabilities using relaxation procedure.
After each iteration, the accuracy of classification is checked by comparing it with the reference file. Such statistics as the errors of omission, errors of commission, the overall accuracy, and the percentage of accuracy improvement are compiled and tabulated in the form of an error matrix. The assessment of algorithms of different strategies of iteration and the compilation of the overall error patterns are performed based on the information from the error matrix and the comparison of the classification map resulting from each iteration with the reference file. If the accuracy is not satisfied, the same step is repeated to update probabilities until the accuracy is converged at a fixed point or starts declining. The whole probabilistic relaxation procedure is graphically illustrated in Figure 2-7. Note that the portion within the dotted frame is the essential part which is different from that of a conventional maximum likelihood classification procedure.

2.5 Summary

The detailed methodology and procedures have been described and their rationale have been explained in this chapter; the main purpose of it is to give a clear procedural picture of the research design, with the help of graphics. The implementing of the procedure is through statistical models which are discussed in the next chapter as well as through the ERDAS programs listed in Table 2-2. The understanding of the geographical relationship among different land cover categories in the study area and the expertise in manual image interpretation is the essential requirement in order to provide ground truth information, as well as the interaction between human and computer perceptions of the land cover categories under study.

Programs to be written include:
a. input/output between disk storage and computer main memory for data manipulation, including data reformatting of the data from ERDAS output;
b. spectral separability checking by divergence using training statistics;
c. spectral separability checking by J-M distance using training statistics;
d. obtaining statistics from the reference map and the original imagery;
e. finding water body boundary pixels for the purpose of error pattern analysis;
f. calculating normalized probability densities of each pixel's being the different land cover categories;
g. conversion of the probability file into a classification map using a maximum likelihood decision rule;
h. printing water body as it is classified after each iteration;
i. compilation and tabulation of error matrix for accuracy check;
j. computation of compatibility coefficients from the output of either programs f or k;
k. updating the probabilities using compatibility coefficients.

Some subroutines are included in the main programs a through h. Their FORTRAN codes are listed in the Appendices.

Table 2-2. Index of the ERDAS programs related to this study.

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALARM</td>
<td>tests signatures by highlighting pixels having the same signature in displayed image file. This program can give a rough idea about the suitability of the selected training statistics.</td>
</tr>
<tr>
<td>CMATRIX</td>
<td>tests homogeneity of each sample within signature set.</td>
</tr>
<tr>
<td>CURSES</td>
<td>uses cursor to interactively display coordinate and pixel information about specific points on a displayed image.</td>
</tr>
<tr>
<td>DIGFLD</td>
<td>extracts and saves training statistics from the existing .DIG files.</td>
</tr>
<tr>
<td>DIGSCRN</td>
<td>digitizes selected areas of an image data file displayed on the screen; creates .DIG file containing polygon, vector and point information with class value.</td>
</tr>
<tr>
<td>Command</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DISPLAY</td>
<td>displays GIS data sets on the image processor display unit.</td>
</tr>
<tr>
<td>DUMPBIL</td>
<td>writes ERDAS image and GIS files to the tape in order to load them into another computer system. This is the main way to transfer data from the SUN workstation to the VAX for data manipulation.</td>
</tr>
<tr>
<td>ELLISPSE</td>
<td>ellipses plot used to test signatures.</td>
</tr>
<tr>
<td>FIELD</td>
<td>uses joystick input to interactively define polygon boundaries for training samples.</td>
</tr>
<tr>
<td>GETSIG</td>
<td>lists content of a signature file.</td>
</tr>
<tr>
<td>GRDPOL</td>
<td>grid-converts .DIG file and inserts the appropriate class values into the corresponding cells of a GIS file. This program is used to create ground truth reference file.</td>
</tr>
<tr>
<td>GISMAP</td>
<td>produces color hard copy maps from GIS files.</td>
</tr>
<tr>
<td>LOADBSQ</td>
<td>copies data from tape in BSQ format.</td>
</tr>
<tr>
<td>READ</td>
<td>displays an image file.</td>
</tr>
<tr>
<td>RGBCLR</td>
<td>modifies interactively the color scheme by assigning intensity levels of red, green, and blue.</td>
</tr>
<tr>
<td>SIGEXT</td>
<td>selects polygon areas from an existing polygon file and use them to extract signatures from the corresponding image file.</td>
</tr>
<tr>
<td>SIGSUB</td>
<td>creates subset of the existing signature file, allowing user to select different band combinations to extract training statistics.</td>
</tr>
<tr>
<td>SUBSET</td>
<td>copies a selected portion and bands of an input data file into an output data file.</td>
</tr>
</tbody>
</table>
CHAPTER III STATISTICAL MODELS

3.1 Multivariate Analysis in Remote Sensing - General

Multivariate analysis is an explanatory data analysis technique used in physical geography in arriving at an understanding of the way in which physical phenomena work. The advantage of system approach in physical geography is that it allows the set of variables under consideration to be viewed as a whole, rather than as the sum of a number of simpler relationships (Chorley and Kennedy 1971). The behavior of a complex interacting system of variables cannot easily or efficiently be described in terms of a number of bivariate relationships. In order to understand such systems, it is necessary to use multivariate analysis.

The essence of multivariate analysis techniques rests on its capability of examining a set of data from different angles and piecing together information about the systems being investigated, which may lead to a subsequent analysis that is more refined and revealing. The premise of this technique is that the relationships between the complicated phenomena in the real world can be represented by a set of measurable variables that are selected carefully and assumed reasonably to have close correlation with the phenomena under investigation. The hypothesis is then tested by inputting the numerical measurements on the variables to statistical models that may lead to an alternative way of explaining phenomena and relationships that geographers are scrutinizing (Mather 1976).

Multispectral classification is a technique developed from multivariate analysis. In multispectral remote sensing, the spectral intensity is closely related to the spectral behavior of actual land cover being imaged. The intensities over different bands are
detected by the sensor and quantified to a digital number that is recorded as gray level on tape. This mode of detecting and recording information has provided the possibility for scientists to investigate the geographical phenomena represented by spectral intensities using computer facilities and statistical pattern recognition techniques. Considering the characteristics of remote sensing data, statistical models have some advantages in practice because:

1. Remote sensing data exhibit many incidental variations which tend to obscure the characteristic differences among the classes of interest due to the inherent randomness of the natural environment. Statistical analysis helps to account for these variations and thus reduces their potentially adverse effects on classification accuracy.

2. It is easy to extract statistics from within the relatively homogeneous areas on remote sensing images that can train a statistical model to establish a discriminating procedure in classification. In some sense, the error rate of classification is controllable and predictable by refining training statistics.

3. In practice, there always exists uncertainty concerning the true identity of the training patterns used to determine the discriminating functions. Statistical methods are tolerant of such errors as long as their frequency of occurrence is relatively low. When modified appropriately, a statistical model can lower the risk of error, or even eliminate ambiguity by incorporating contextual information into classification.

4. The pattern classes of interest may actually overlap in the measurement space, i.e., some of measurements from one class may be indistinguishable from those made to other classes. In such situations, statistical pattern recognition methods allow classifications which are "most often" or "most probably" correct.
3.2 Statistical Characterization of Spectral Intensities - Mean Vector and Covariance Matrix

In multivariate analysis, the measurements on n variables made to m samples are projected to an n-dimensional space. The position of a sample in an n-dimensional space is determined by the measurements and the statistical characteristics of these measurements are represented by mean vector and an n by n covariance matrix. The elements of this matrix consist of numbers that describe the variance of measurements between the same variables and covariance of measurements made between different variables, and so it appears to be a symmetric matrix. The variance is indicative of the distribution of measurement on a variable around its mean value, the same indication of standard deviation, but in a different scale. The covariance, in a sense, is a measure of the relationship, or association, between the variables $X_i$ and $X_j$. When there is a high probability that large values of $X_i$ go with large values of $X_j$, the covariance between $X_i$ and $X_j$ is positive; and when there is a high probability that large values of $X_i$ go with small values of $X_j$, or vice versa, the covariance between $X_i$ and $X_j$ is negative. When $X_i$ and $X_j$ are independent from each other, the covariance is zero (Freund 1971). The magnitude of covariance matrix elements reflects the numerical dispersion of measurements on n variables and the degree of positive or negative correlation between different variables.

In the remote sensing practice, the gray levels recorded over n different bands of sensor are analogue to the measurements on n variables. Those pixels within a training field outlined by a polygon for a class pattern are samples for that particular class. More specifically, in a one dimensional, or one band situation, let $\mu_i$ be the estimated

* Note: for all mathematical symbols, see Table 3-3 Notional Index of Section 3.9 in page 75.
mean value of that band for class $i$ and $\beta_i^2$ be variance for class $i$, then $\mu_i$ and $\beta_i^2$ are given:

$$\mu_i = \frac{1}{q_i} \sum_{k=1}^{q_i} x_k \quad (3-1)$$

$$\beta_i^2 = \frac{1}{q_i} \sum_{k=1}^{q_i} (x_k - \mu_i)^2 \quad (3-2)$$

where $i=1, 2, ..., m$ is the number of category, $q_i$ is the number of training pixels in class $i$, and $X_j$ is the gray level of the $j$th pixel within the training area for class $i$.

![Figure 3-1. N-band training data for class $i$, where $x_{jk}$ is the gray level on band $j$ of the $k$th pixel for class $i$.](image)

In multispectral remote sensing, the statistical pattern for a particular class is characterized using training pixels (Figure 3-1). Let $\mu_{ij}$ be the mean of gray levels over band $j$ for class $i$, $\beta_{ijl}$ be covariance between band $j$ and $l$ for class $i$, the $\mu_{ij}$ and $\beta_{ijl}$ are given:

$$\mu_{ij} = \frac{1}{q_i} \sum_{k=1}^{q_i} x_{jk} \quad j=1, 2, ..., n; \quad (3-3)$$
\[
\beta_{ijj} = \frac{1}{q_i - 1} \sum_{k=1}^{q_i} (x_{jk} - \mu_{ij})^2 \quad \text{when } j = l, \quad (3-4)
\]

and
\[
\beta_{ijl} = \frac{1}{q_i - 1} \sum_{k=1}^{q_i} (x_{jk} - \mu_{ij})(x_{lk} - \mu_{il}) \quad \text{when } j \neq l \text{ and } j, l = 1, 2, \ldots, n. \quad (3-5)
\]

For the general situation of multispectral remote sensing, the representation of the statistical characterization for an \(n\)-band data set is given in matrix format for the notational convenience:

\[
U_i = \begin{bmatrix}
\mu_{i1} \\
\mu_{i2} \\
\vdots \\
\mu_{ij} \\
\vdots \\
\mu_{in}
\end{bmatrix}, \quad \Sigma_i = \begin{bmatrix}
\beta_{i11} & \beta_{i12} & \cdots & \beta_{i1j} & \cdots & \beta_{i11} & \cdots & \beta_{i1n} \\
\beta_{i21} & \beta_{i22} & \cdots & \beta_{i2j} & \cdots & \beta_{i21} & \cdots & \beta_{i2n} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\beta_{ij1} & \beta_{ij2} & \cdots & \beta_{ijj} & \cdots & \beta_{ij1} & \cdots & \beta_{ijn} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\beta_{i11} & \beta_{i12} & \cdots & \beta_{i1j} & \cdots & \beta_{i11} & \cdots & \beta_{i1n} \\
\beta_{i21} & \beta_{i22} & \cdots & \beta_{i2j} & \cdots & \beta_{i21} & \cdots & \beta_{i2n} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
\beta_{in1} & \beta_{in2} & \cdots & \beta_{inn}
\end{bmatrix} \quad (3-6)
\]

where \(U_i\) is the mean vector for class \(i\) with matrix element \(\mu_{ij}\) as denoted by Equation (3-3) and \(\Sigma_i\) is the covariance matrix for class \(i\) with matrix element \(\beta_{ijl}\) as denoted by Equation (3-5). The mean vectors and covariance matrices for all of \(m\) classes are estimated from training data. This estimation is then used by statistical model to establish the classification scheme.
3.3 Statistical Characterization of Class Membership - Probability Density Function

Given \( n \) digital numbers of a pixel over \( n \) bands, the spectral pattern of this pixel in an \( n \)-dimensional space is determined. This pattern is compared with all numerical spectral patterns for classes summarized in the form of a mean vector and covariance matrix; its resemblance, or closeness, to any of these class patterns is approximated by probability density function associated with each class, provided that the multispectral measurements of all pixels give a normalized distribution, which has been observed repeatedly in most remote sensing data and applications. For the one band, or univariate case, let \( p(x|\mathbf{w}_i) \) be the probability of pixel \( x \) to be class \( i \), the probability density function for that pixel is given:

\[
p(x|\mathbf{w}_i) = \frac{1}{\sqrt{2\pi \beta_i}} \exp\left[-\frac{1}{2} \frac{(x - \mu_i)^2}{\beta_i^2}\right]
\]

(3-7)

where expression \( \exp[ \ ] \) refers to the \( e \) (the base of natural logarithms) raised to the indicated power in bracket;

\( \mu_i = E[x|\mathbf{w}_i] \) is the mean of gray levels from training data for class \( i \);

\( \beta_i^2 = E[(x-\mu_i)^2] \) is the variance of gray levels from training data for class \( i \).

In the case of two bands, or bivariate measurement, the probability density function for each class is estimated by tabulating the frequencies of occurrence for all possible pairs of data values, each pair consisting of a gray level \( x_1 \) from band 1 and gray level \( x_2 \) from band 2. Let \( p(x_1, x_2|\mathbf{w}_i) \) be the probability of pixel \( x \), with spectral measurements over two bands, to be class \( i \), the bivariate probability density function is given by:
\[ p(x_1, x_2 | w_i) = \frac{1}{2\pi(\beta_{i11}\beta_{i22} - \beta_{i12}^2)^{1/2}} \times \exp \left[ -\frac{1}{2} \left( \frac{(x_1 - \mu_{i1})^2}{\beta_{i11}} - \frac{2\beta_{i12}(x_1 - \mu_{i1})(x_2 - \mu_{i2})}{\beta_{i11}\beta_{i12}} + \frac{(x_2 - \mu_{i2})^2}{\beta_{i22}} \right) \right] \]

where \( \mu_{i1} = \mathbb{E}[x_1 | w_i] \) is the mean of gray levels from training data in band 1 for class \( i \); 
\( \mu_{i2} = \mathbb{E}[x_2 | w_i] \) is the mean of gray levels for training data in band 2 for class \( i \); 
\( \beta_{i11} = \mathbb{E}[(x_j - \mu_{i1})^2] \) is the variance of gray levels from training data in band 1 for class \( i \); 
\( \beta_{i22} = \mathbb{E}[(x_j - \mu_{i2})^2] \) is the variance of gray levels from training data in band 2 for class \( i \); 
\( \beta_{i12} = \beta_{i21} = \mathbb{E}[(x_1 - \mu_{i1})(x_2 - \mu_{i2})] \) is the covariance of gray levels from training data between band 1 and band 2 for class \( i \).

Conceptually, we can continue this approach to multispectral data to include more bands up to \( n \). Considering the complexity of Equation (3-8), the equation for the \( n \)-band situation becomes quite complicated. Again, for the notational convenience, the general case of \( n \)-band data is expressed in a compact way using matrix notation as:

\[
X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_l \\ \vdots \\ x_n \end{bmatrix}, \quad U_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ij} \\ \vdots \\ \mu_{il} \\ \vdots \\ \mu_{in} \end{bmatrix}, \quad \Sigma_i = \begin{bmatrix} \beta_{i11} & \beta_{i12} & \cdots & \beta_{i1j} & \cdots & \beta_{i1l} & \cdots & \beta_{i1n} \\ \beta_{i21} & \beta_{i22} & \cdots & \beta_{i2j} & \cdots & \beta_{i2l} & \cdots & \beta_{i2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_{ij1} & \beta_{ij2} & \cdots & \beta_{ijj} & \cdots & \beta_{ijl} & \cdots & \beta_{ijn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_{il1} & \beta_{il2} & \cdots & \beta_{ilj} & \cdots & \beta_{ill} & \cdots & \beta_{iln} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_{in1} & \beta_{in2} & \cdots & \beta_{ijn} & \cdots & \beta_{inl} & \cdots & \beta_{inn} \end{bmatrix} \quad \text{(3-9)}
\]
where $\mu_{ij} = E[X_j|W_i]$ is the mean of gray levels for training data in band $j$ for class $i$;

$\sigma_{ij}^2 = E[(X_j - \mu_{ij})^2]$ is the variance of gray levels for training data in band $j$ for class $i$;

$\beta_{ij} = E[(X_1 - \mu_{i1})(X_2 - \mu_{i2})]$ is the covariance of gray levels from training data between band $j$ and band $l$ for class $i$.

Let $X$ be data vector, $U_i$ be the mean vector for class $i$, and $\Sigma_i$ be the covariance matrix for class $i$, $p(X|W_i)$ be the probability density value for the measurement vector $X$ to be class $i$, then the $n$ band probability density function is given:

$$p(X|W_i) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma_i|}} \exp\left[-\frac{1}{2}(X - U_i)^T\Sigma_i^{-1}(X - U_i)\right]$$  \hspace{1cm} (3.10)

where $|\Sigma_i|$ is the determinant of the covariance matrix $\Sigma_i$, $\Sigma_i^{-1}$ is the inverse of $\Sigma_i$, and $(X - U_i)^T$ is the transpose of the vector $(X - U_i)$. Now we can denote the normal density function $p(X|W_i)$ with mean vector $U_i$ and covariance matrix $\Sigma_i$ by writing

$$p(X|w_i) \sim \mathcal{N}(U_i, \Sigma_i)$$  \hspace{1cm} (3.11)

for the notational convenience.

### 3.4 Determination of Class membership - Maximum Likelihood Decision Rule.

A discriminating function is the one that is able to divide an $n$-dimensional space into separate decision regions, each region corresponding to a specific discriminable class (Gendern and Uiterwijk 1987). The construction of a classifier should be such that it can identify any measurement vector's belonging to the class corresponding to the
decision region in which it falls. A relatively simple two dimensional case is conceptually illustrated in Figure 3-2 that three classes are separated exclusively by selecting appropriate decision boundaries represented by discriminating functions.

Considering the complexity of natural environment, the situation is never so simple like that of Figure 3-2. In multispectral remote sensing, the most common situation is the overlap of the decision regions, as is the case between residential and grass in an urban scene shown in Figure 1-3. Now that after essential statistics related to our remote sensing problem have been introduced based on the assumption of normal distribution of remote sensing data, the maximum likelihood decision rule is applied as a set of discriminating functions to solve the class discrimination problem. Using the same notation for Equation (3-10), the maximum likelihood decision rule is formally stated as:

\[
\text{Decide } X \text{ belongs to } W_i, \text{ if and only if } \\
p(X|w_i)p(w_i) \geq p(X|w_j)p(w_j) \quad (3.12)
\]

for all \( j=1, 2, ..., i-1, i+1, ..., m \)
where \( p(W_i) \) is the \textit{a priori} probability associated with class \( i \), or the probability of observing the pattern of class \( i \) independent of any other information. That is, the machine computes the products of \( p(X|W_i)p(W_i) \) for all \( i \) and assigns the pixel to the class which has the maximum product. This procedure is schematically illustrated in Figure 3-3.

Here the concept of \textit{a priori} probability \( p(W_i) \) is worth of discussion because its selection may have some essential influence on the actual assignment of class membership. If the \textit{a priori} probabilities for all classes are not chosen appropriately, it may deteriorate the accuracy of classification. To make this point more clearly, let us consider a simple numerical example of a two class situation using equations defined so far. Let two band measurement vector \( X \) be \((4,3)\), \( U_1=(4,2) \), \( U_2=(3,3) \) be mean vectors from training patterns for class 1 and class 2 respectively, and
\[
\Sigma_1 = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 4 & 5 \\ 5 & 7 \end{bmatrix} \quad \Sigma_1^{-1} = \begin{bmatrix} 3 & -2 \\ -2 & \frac{3}{2} \end{bmatrix} \quad \Sigma_2^{-1} = \begin{bmatrix} \frac{7}{3} & -\frac{5}{3} \\ -\frac{5}{3} & \frac{4}{3} \end{bmatrix}
\]

be covariance matrices and their corresponding inverse matrices for class 1 and class 2 respectively. \( |\Sigma_1| = 2 \) and \( |\Sigma_2| = 3 \) be determinants of \( \Sigma_1 \) and \( \Sigma_2 \), then the quadratic products of Equation (3-10) for class 1 is given:

\[
(X - U_1)^T \Sigma_1^{-1} (X - U_1) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{3}{2} = 1.5
\]

The probability density value for class 1 is then:

\[
p(X|w_1) = \frac{1}{\sqrt{2\pi|\Sigma_1|}} (X - U_1)^T \Sigma_1^{-1} (X - U_1) = \frac{e^{-\frac{1}{2} \times \frac{3}{2}}}{\sqrt{2\pi \times 2}} = 0.0532
\]

Similarly for the second class,

\[
(X - U_2)^T \Sigma_2^{-1} (X - U_2) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{7}{3} & -\frac{5}{3} \\ -\frac{5}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{2}{3} = 1.667
\]

and

\[
p(X|w_2) = \frac{1}{\sqrt{2\pi|\Sigma_2|}} (X - U_2)^T \Sigma_2^{-1} (X - U_2) = \frac{e^{-\frac{1}{2} \times \frac{2}{3}}}{\sqrt{2\pi \times 3}} = 0.0338
\]

Thus, the measurement vector \( (4, 3) \) has a higher probability associated with the membership in class 1 than with that in class 2, and it is appropriate to classify \( X \) into class 1.
The calculations above assume equal \textit{a priori} probability for class membership for both class 1 and class 2. Removing this restriction, we take \textit{a priori} probability into account. Assume that the $p(w_1)=1/3$ and $p(w_2)=2/3$ are observed, then we calculate the probability for two classes as:

$$p(X|w_1)p(w_1) = 0.0532 \times \frac{1}{3} = 0.0177$$

and

$$p(X|w_2)p(w_2) = 0.0338 \times \frac{2}{3} = 0.0225$$

Thus, the class 2 is favored for the observation (4 3) over class 1, i.e., the class membership in the first situation when the \textit{a priori} probability is not considered is switched to class 2 when \textit{a priori} probability is taken into account.

The setting of \textit{a priori} probability is a complicated topic differing in its complexity under different contexts, which is beyond the discussion of this study. In remote sensing application, however, a common practice is to assign the \textit{a priori} probability to a class according to the physical dimension this particular class is assumed to occupy. In our numerical example, the class 1 may occupy one third of the area represented by the entire image and class 2 occupy the other two thirds. The advantage of this approach is that it meets the requirement that the conditional probabilities for all classes sum to 1, and it is easy to understand conceptually. The disadvantage of it is that setting a very small \textit{a priori} probability for a class in terms of its area occurrence may effectively remove this class from output classification. Considering an image of arid area consisting of 512 by 512 pixels, with only a small area of open water consisting of 400 pixels (in TM imagery, these pixels may represent, say, a small lake of 0.14 square miles). The \textit{a priori} probability in terms of area occurrence of water is only 0.0015.
With so low \textit{a priori} probability of occurrence, the water class may be actually removed during classification unless the probability density value for water is substantially greater than any other classes.

A further detailed scrutiny to the area occurrence assumption for remote sensing application reveals, however, that it is not so valid in statistical theory. In fact, the \textit{a priori} probability consists of two important parts: $p(w_i|X)$, the probability that an observation is drawn from all classes in terms of area occurrence when the measurement vector is $X$, and the $p(X)$, the probability that the measurement vector $X$ is observed from all spectral patterns, or using the term defined for this study, the occurrence of a specific spectral pattern. The Law of Conditional Probability states that:

$$p(w_i) = \frac{p(w_i|X)}{p(X)}$$

(3-13)

The Equation (3-13) states exactly the fact that the occurrence of class $i$ among all classes is closely associated with the occurrence of spectral pattern for class $i$, when it is otherwise independent of any other information. In our previous example of arid area, when the occurrence of class $i$ is only 400 among 262,144 pixels (512 by 512), its occurrence of the associated spectral pattern is close to the same probability, i.e., the $p(w_i)\approx 1$. The same situation is also applicable to all other classes.

Summarizing the above discussion, it is concluded that the class occurrence goes with the occurrence of its associated spectral pattern. When we have no reason to favor one class over any other classes in the class membership decision stage for our urban problem, like that discussed in Chapter 1, it is reasonable to assume that the \textit{a priori} probabilities for all classes are equal to one. In other words, the occurrence of classes...
in our particular urban situation is associated only with the occurrence of specific spectral patterns.

In practice, the product in the bracket of Equation (3-10) tends to reach the magnitude of one hundred that makes it difficult to compare the magnitude of \( p(X|W_i) \) and \( p(X|W_j) \) because both of them will approach zero and thus not comparable. If we applied the reduction factor to the quadratic product of Equation (3-10), we can not guarantee that the \( p(X|W_i) \) and \( p(X|W_j) \) keep the same relationship. To make this point more clear, let

\[
p(X|W_i) = \frac{1}{|\Sigma_i|} e^{-\beta_i}
\]

be the simpler form of Equation (3-10). When \( p(X|W_i) \) is greater than \( p(X|W_j) \), we have

\[
\frac{1}{|\Sigma_i|} e^{-\beta_i} > \frac{1}{|\Sigma_j|} e^{-\beta_j}
\]

(3-15)

Note that both \( \Sigma_i \) and \( \beta_i \) have influence on the magnitude of \( p(X|W_i) \). Suppose that \( |\Sigma_i| \) and \( |\Sigma_j| \) are 100 and 00.1, \( \beta_i \) and \( \beta_j \) are 100 and 200, respectively, then

\[
\frac{1}{100} e^{-100} > \frac{1}{0.01} e^{-200}
\]

is true. When a reduction factor of 0.01 is applied to both \( \beta_i \) and \( \beta_j \) without considering the possible influence of \( \Sigma_i \) and \( \Sigma_j \) on the \( p(X|W_i) \) and \( p(X|W_j) \), the relationship

\[
\frac{1}{100} e^{-0.01} > \frac{1}{0.01} e^{-0.02}
\]
is no longer true.

To circumvent this difficulty, a logarithmic transformation is applied to the both side of Equation (3-10) that makes $p(X|w_j)$ and $p(X|w_j)$ comparable and is given

$$\log_e[p(X|w_j)] = -\frac{1}{2}\log_e[2\pi] - \frac{1}{2}\log_e|\Sigma_i| - \frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) \quad (3.16)$$

When the reduction factor of 0.01 is applied to the right side of Equation (3-16) and the reverse operation is performed to the logarithmic term of Equation (3-16), if

$$\log_e[p(X|w_j)] = -\beta_i > \log_e[p(X|w_j)] = -\beta_j$$

is true, then

$$p(X|w_i) = e^{-0.01\beta_i} > p(X|w_j) = e^{-0.01\beta_j}$$

is also true.

The Equation (3-16) is the standard formula for the calculation of logarithmic-probabilities. After the application of reduction factor, it is converted to the class membership probabilities through the reversal of the logarithmic operation.

### 3.5 Statistical Separability - Divergence and J-M Distance

In the problems of feature extraction, the optimum selection of features based on a set of measurements is that minimizes the probability of error. However, in most cases, direct minimization of probability error as to determine an optimal feature is impossible in the context of statistical theory, because the meaningful functional
expression for error probability is too difficult to find. Therefore, it is useful to search a feature selection criterion that is weaker than that of error probability, but easier to evaluate and manipulate (Kailath 1967). In general, the statistical distance between two probability distributions is quite a useful criterion to determine the separability of two features in an n-dimensional space. The basic assumption for separability checking using statistical distance is that the farther apart these two features are from each other, the lower the risk of mistaking one feature for another in the following classification. So, various distance measures have been studied as simpler substitutes for the estimate of error probability as the evaluation of the optimum feature set selection (Brooner, Haralick and Dinstein 1971).

What we would like to search for such distance measures is a property of such type that when the distance between two distributions for a feature i is greater than that of feature j, then the error probability in the classification for feature i is always less than that for feature j. In the engineering literature, the divergence and J-M distance are two widely accepted statistical distance measures applied to the problems of optimal signal selection (Sebestyeu 1962). In a remote sensing context, the reflectance or emittance from land cover is received by sensors as "signal" and quantified as "digital numbers" that is the representation of this land cover category on remote sensing imagery. The normal distribution of the digital numbers and the spectral signature of different land cover categories makes it possible to compare the statistical separability between the pairs of land cover categories using divergence and J-M distance. Theoretically speaking, the farther the selected features in an n-dimensional measurement space in terms of statistical distances, the better the result of the classification, because only those features that are spectrally separable tested by means of statistical distance are used in classification. Here, the "better" generally means the lower risk of classification error.
The divergence is a measure of difficulty of discriminating between two distributions i and j. By definition, the divergence is the difference in the mean values of the logarithmic-likelihood ratio of the two features and given by:

\[ D_{ij} = E_i[\log_e(L(X))] - E_j[\log_e(L(X))] \tag{3-17} \]

where \( E[ ] \) denotes the mean value, \( \log_e(L(X)) \) denotes the logarithmic-likelihood ratio of the feature i and j with measurement \( f_i(x_1, x_2, ..., x_k, ..., x_n) \) and \( f_j(x_1, x_2, ..., x_k, ..., x_n) \) while

\[ \log_e(L(x)) = \log_e \left[ \frac{f_i(x_1, x_2, ..., x_k, ..., x_n)}{f_j(x_1, x_2, ..., x_k, ..., x_n)} \right] \tag{3-18} \]

Thus, the general form of divergence is given:

\[ D_{ij} = \int_{x_1, x_2, ..., x_k, ..., x_n} \left[ f_i(x_1, x_2, ..., x_k, ..., x_n) - f_j(x_1, x_2, ..., x_k, ..., x_n) \right] \times \log_e \left[ \frac{f_i(x_1, x_2, ..., x_k, ..., x_n)}{f_j(x_1, x_2, ..., x_k, ..., x_n)} \right] \, dx_1 \, dx_2 \, ... \, dx_k \, ... \, dx_n \tag{3-19} \]

More specifically in the problems of remote sensing application, the divergence between pairs of features with probability density functions is:

\[ D_{ij} = \int_x \left[ p(x|w_i) - p(x|w_j) \right] \log_e \left[ \frac{p(x|w_i)}{p(x|w_j)} \right] \, dx \tag{3-20} \]
Under the assumption of normal probability density function for all features, the Equation (3-20) becomes the expression that has removed the integrals and involves only the mean vector and covariance matrix, making it mathematically more tractable:

\[
D_{ij} = \frac{1}{2} \text{tr} \left[ \left( \Sigma_i - \Sigma_j \right) \left( \Sigma_j^{-1} - \Sigma_i^{-1} \right) \right] + \frac{1}{2} \text{tr} \left[ \left( \Sigma_i^{-1} + \Sigma_j^{-1} \right) (U_i - U_j)(U_i - U_j)^T \right]
\]  

(3-21)

where \( \text{tr}[A] \) denotes the trace, or the sum of diagonal elements of square matrix \( A \).

Some mathematical properties of divergence have been noted. We should be cautious when taking these properties into account of remote sensing image classification problems, though these properties have been proved by Kullback (1959):

1) \( D_{ij} > 0 \). The divergence is always greater than zero.

This property is apparent from Equation (3-19) because it involves the volume integrals over the entire measurement space in the theoretical expression of divergence. In the remote sensing practice related to Equation (3-21), however, it is not necessarily always true; in its expression, the divergence is associated only with the mean vector and covariance matrix, which in turn, reflect the degree of suitability of the selected training data. In general, the carefully selected training data in correspondence with the criteria listed in Chapter 2, Section 2.2 always keeps the divergence positive; otherwise, we may have reason to question the validity of the selected training statistics on whether they can be used to calculate divergence between class \( i \) and \( j \) as the means to check the spectral separability.

2) \( D_{ii} = 0 \). The divergence of a probability density function related to itself is zero.
This is also apparent from Equation (3-21) because for the same training data, the 
$\sum_{i} = \sum_{j}$ makes the first term of this expression zero and $U_{i} = U_{j}$ makes the second 
term zero.

3) $D_{ij} = D_{ji}$. The symmetricity of the divergence.

In the general case of divergence expressed by Equation (3-19), note that

$$
\log_{e} \left[ \frac{f_{i}(x_{1}, x_{2}, \ldots, x_{k}, \ldots, x_{n})}{f_{j}(x_{1}, x_{2}, \ldots, x_{k}, \ldots, x_{n})} \right] = - \log_{e} \left[ \frac{f_{j}(x_{1}, x_{2}, \ldots, x_{k}, \ldots, x_{n})}{f_{i}(x_{1}, x_{2}, \ldots, x_{k}, \ldots, x_{n})} \right]
$$

(3-22)

makes this property apparent when we apply Equation (3-22) to Equation (3-19).

4) Divergence is additive if the components of measurement vector are statistical 
independent, that is:

$$
D_{ij}(x_{1}, x_{2}, \ldots, x_{n}) = \sum_{k=1}^{n} D_{ij}(x_{k})
$$

(3-23)

Again, for the remote sensing application, it is not always true because the 
measurements on some bands have close correlation like that between band 2 and 
band 3 of MSS and TM imagery.

5) Adding new measurements to a set of classes never decreases the statistical 
separability, or:

$$
D_{ij}(x_{1}, x_{2}, \ldots, x_{k}, \ldots, x_{n}, x_{n+1}) \geq D_{ij}(x_{1}, x_{2}, \ldots, x_{n})
$$

(3-24)

This property is closely related with the property 1). In remote sensing practice, 
its performance may vary with the degree of independency of measurements between 
different bands.
The J-M distance, like the divergence, is another intuitive way to measure statistical separability of pairs of classes. Roughly speaking, it measures the average difference between the two class density functions as the evaluation of feature separation. By definition, the general form of J-M distance is given:

$$J_{ij} = \left\{ \int \left[ \frac{\sqrt{f_i(x_1, x_2, \ldots, x_k, \ldots, x_n)}}{\sqrt{f_j(x_1, x_2, \ldots, x_k, \ldots, x_n)}} \right]^2 \right\}^{1/2} dx_1 dx_2 \ldots dx_k \ldots dx_n$$  \hspace{1cm} (3-25)

When probability density functions for class $i$ and class $j$ are given, it becomes:

$$J_{ij} = \left\{ \int \left[ \sqrt{p(x|w_i)} - \sqrt{p(x|w_j)} \right]^2 \right\}^{1/2}$$  \hspace{1cm} (3-26)

Again, under the assumption of normal distribution, the Equation (3-26) is reduced to

$$J_{ij} = \left[ 2 \left( 1 - e^{-\beta} \right) \right]^{1/2}$$  \hspace{1cm} (3-27)

where,

$$\beta = \frac{1}{8} (U_i - U_j)^T \left( \frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (U_i - U_j) + \frac{1}{2} \ln \left\{ \frac{\left( \Sigma_i + \Sigma_j \right)/2}{\left[ \Sigma_i \times \Sigma_j \right]^{1/2}} \right\}$$  \hspace{1cm} (3-28)

which involves only the mean vector and covariance matrix but no integrals.

In practice, it is often desirable to apply a reduction coefficient to $\beta$ and then a magnification coefficient to $J_{ij}$ in order to stretch the contrast of J-M distances between pairs of classes. Suppose the coefficient $\beta_{ij}$ for class $i$ and $j$ and $\beta_{ik}$ for class $i$ and $k$ have a relatively larger number, say, in the scale of three digits as, for example,
\( \beta_{ij} = 125 \) and \( \beta_{ik} = 625 \). Then the corresponding \( J_{ij} \) and \( J_{ik} \) both approach very close to 2 which makes it impossible to compare how far class i is from j with that of class i from k. When a reduction coefficient 0.001 is applied to both \( \beta_{ij} \) and \( \beta_{ik} \), then \( \beta_{ij} = 0.125 \) and \( \beta_{ik} = 0.625 \) makes a substantial difference between \( J_{ij} \) and \( J_{ik} \). This difference can be crucial to make a decision on whether to merge class j with class i or keep it a separate class from class i. Note that from Equation (3.26) if \( \beta_{ij} < \beta_{ik} \), then \( J_{ij} < J_{ik} \); if the same reduction coefficient \( c \) is applied to both \( \beta_{ij} \) and \( \beta_{ik} \), the relationship of \( J_{ij} < J_{ik} \) is still held because that the \( c\beta_{ij} < c\beta_{ik} \) holds.

Figure 3-4. Divergence and J-M distance as functions of normalized distance.
The J-M distance has all of the properties listed for the divergence. However, an apparent advantage of J-M distance over divergence is its "saturated" behavior arising from the negative exponential term of Equation (3-27). The effect of this term is to give an exponentially decreasing weight to increasing differences between the class density functions until the Jij approaches a certain limit. This fact is more easily to be understood graphically in Figure 3-4. In Figure 3-4a, the divergence as the function of normalized distance approaches to ∞, which is hard to understand in its physical meaning. In Figure 3-4b on the other hand, the feature separability as function of the normalized distance approaches to a certain limit when the probability of error drops to zero, exactly what we expect to achieve by checking the spectral separability using training statistics.

3.6 Statistical Characterization of the Contextual Information - Compatibility Coefficient

Suppose that we are analyzing an image, with the aim of classifying it, and we have detected a set of land cover categories \( W = (W_1, W_2, ..., W_k, ..., W_n) \) on the image that they have higher spectral separability from one another. During the initial maximum likelihood classification, the identification of certain percentage of pixels is still not determined, or is ambiguous, since each pixel is classified independently of one another without considering the relationships that exist among them. Our purpose now is to design a statistical model that characterizes the relationships existing among them and represent them by means of compatibility coefficients. In Sections 3.6 and 3.7, we are not going to deal with image file any more. Instead, we will deal with \( m \) probability files, each of them is an \( n \) by \( n \) matrix, to establish the probabilistic relaxation
procedure. Here, m denotes the actual number of the category to be assigned by the classification scheme.

Before defining compatibility coefficient, let us consider the desirable characteristics of relaxation operator R should have when it is operated on the probabilities of class membership over the entire probability file. The probability \( p_i(w_k) \) of a given category \( w_k \) for a given pixel at the location \( i \) is increased by \( R \) if the categories of other pixels around the location \( i \) that have higher probability are highly compatible with \( w_k \) at the location \( i \). Conversely, the \( p_i(w_k) \) is decreased by \( R \) if other high probability categories for pixels around the location \( i \) are incompatible with the \( w_k \) at the location \( i \). On the other hand, categories having low probabilities have little influence on \( p_i(w_k) \), whether or not they are compatible with it. These characteristics are summarized in tabular form as a so-called R operator in Figure 3-5, where "+" means the \( p_i(w_k) \) increases, "-" means that it decreases, and "0" means that it remains relatively unchanged.

<table>
<thead>
<tr>
<th>Compatibility of ( W_j ) with ( W_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High</strong></td>
</tr>
<tr>
<td>( p_i(W_j) ) (High)</td>
</tr>
<tr>
<td>( p_i(W_j) ) (Low)</td>
</tr>
</tbody>
</table>

where \( j, k=1, 2, ..., m \)

Figure 3-5. The behavior of the R operator to be used in updating probabilities.

Now we come closer to the desired behavior of operator R by making use of compatibility coefficients that take on both positive and negative values in the range [---]
Such coefficients are regarded as representing "mutual information" of the categories at the neighboring pixels. Let us denote compatibility coefficient of category $W_j$ at the location $i$ with category $W_k$ at the specific neighbor of pixel $i$ by $R_{i,\partial}(W_j, W_k)$. Note that $i$ takes values of all location over the entire probability file, including the location at corners and edges while $\partial$ takes values 1 through 9, denoting that we take the category probabilities of 8 neighboring pixels plus that of the central pixel itself (at location 5 in Figure 3-5) into account of compatibility coefficient calculation. We would like that the $R$'s behave as follows:

a) if $W_j$ for pixel $i$ frequently co-occurs with $W_k$ for pixel at the neighbor $\partial$, then $R_{i,\partial}(W_j, W_k) > 0$;

b) if $W_j$ for pixel $i$ rarely co-occurs with $W_k$ for pixel at the neighbor $\partial$, then $R_{i,\partial}(W_j, W_k) < 0$;

c) if neither category is constrained by the other, then $R_{i,\partial}(W_j, W_k) = 0$, i.e., they are independent of each other.

d) the magnitude of $R_{i,\partial}(W_j, W_k)$ represents the strength of compatibility.

Before starting the construction of compatibility coefficients and relaxation, a prerequisite is to normalize the probability density values, a theoretical requirement for the whole probabilistic relaxation procedure. In fact, when we assume the use of a priori probabilities in the Section 3.5, we have also made the assumption that the probabilities we are dealing with are conditional probabilities instead of only probability density values, even if we have assumed the equal a priori probability for all classes. Let $p_i(W_k)$ be normalized probability of pixel $i$ as category $k$, then

$$p_i(W_k) = \frac{p(x|W_k)}{\sum_{k=1}^{m} p(x|W_k)}$$

(3-29)
Note that the normalization does not change the result of classification by maximum likelihood decision rule because if \( p(x|w_k) > p(x|w_j) \) for all \( j \) holds,

\[
p_i(w_k) = \frac{p(x|w_k)}{\sum_{k=1}^{m} p(x|w_k)} > p_i(w_j) = \frac{p(x|w_j)}{\sum_{k=1}^{m} p(x|w_j)}, \quad j = 1, 2, ..., m
\]

still holds.

There are several alternatives to calculate the compatibility coefficients. The one adopted in this approach is suggested by Peleg and Rosenfeld (1978) because it gives the emphasis on the mutual information among neighboring pixels.

We estimate the probability of any pixel's being the category \( k \) by:

\[
\bar{p}(w_k) = \frac{1}{n} \sum_{i=1}^{n} p_i(w_k)
\]

and the joint probability of a pair of pixels with pixel \( i \) as being the category \( j \) and the pixel at the neighbor \( \partial \) of the pixel \( i \) as being the category \( k \) by:

\[
p_{i,\partial}(w_j, w_k) = \frac{1}{n} \sum_{k=1}^{m} p_i(w_j) p_{i,\partial}(w_k)
\]

where \( p_i(w_j) \) is the initial estimate of probability as category \( j \) for pixel \( i \) and \( (i, \partial) \) denote the specific neighbor of pixel \( i \). Now we estimate the conditional probability that pixel \( i \) is category \( w_j \) given the pixel \( (i, \partial) \) is category \( w_k \) by:

\[
p_{i,\partial}(w_j|w_k) = \frac{p_{i,\partial}(w_j, w_k)}{\bar{p}(w_k)} = \frac{\sum_{k=1}^{n} p_i(w_j) p_{i,\partial}(w_k)}{\sum_{k=1}^{m} p_i(w_k)}
\]
For any event A, whose probability of occurrence is \( P(A) \), the amount of information we receive as a result of being told that A has occurred is defined as \( I(A) = -\log_e P(A) \). In the same manner, the conditional information that we receive when we already know that B has occurred is \( I(A|B) = -\log_e P(A|B) \), provided that A has occurred. The contribution of B to the information A is expressed by the mutual information

\[
I(A, B) = I(A) - I(A|B) = \log_e \frac{P(A|B)}{P(A)}
\]  

(3-34)

Note that when A is highly correlated with B, \( P(A|B) \) is close to one, making \( I(A|B) \) close to zero and \( I(A, B) \) high; while if A is negatively correlated with B, \( P(A|B) \) will substantially less than one (but greater than zero), making \( I(A|B) \) positive and \( I(A, B) \) small, or even negative. This is exactly the way we would like the compatibility coefficients to behave.

Using Equations (3-31), (3-33) and (3-34), the compatibility coefficient is defined as:

\[
R_{i,j} (w_j, w_k) = \frac{1}{m} \log_e \frac{\sum_{k=1}^{n} p_i(w_j) p_{i,j} (w_k)}{\sum_{k=1}^{n} p_i(w_j) \sum_{k=1}^{n} p_i(w_k)}, \text{ where } j, k = 1, 2, \ldots, m
\]  

(3-35)

The purpose of dividing the result of logarithmic operation by the number of categories is to obtain the compatibility coefficients \( R_{i,j} (w_j, w_k) \) in the range of \([-1, 1]\).

By the configuration of Equation (3-35), we know that the compatibility coefficient matrix is \( m \) by \( m \) by 9 in dimension. The numerator of the Equation (3-35) is not so straightforward and let us discuss it a little bit more on how it proceeds in the construction of the coefficients by Figure 3-6.
Figure 3-6. The conversion of the probabilities to compatibility coefficients.

In fact, the algorithm that creates the compatibility coefficients continuously aggregates the contribution, either positive or negative, of the mutual information to the statistical compatibility among categories summarized over m probability files. Each of the files is a probability matrix with n by n real number elements that is first created by the GML classification and then the probabilities are updated by each iteration. Particular, let us suppose that \( w_2 \) be the probability file for the category labeled 2 and \( w_4 \) be the probability file for the category labeled 4, then the example given in Figure 3-6 shows how the mutual information at the neighbor 3 (\( \partial=3 \)) of the central pixel located at the position 5 (\( \partial=5 \)) is converted to the compatibility coefficient on computer as the element (2, 4, 3) of the compatibility coefficient matrix. The product of \( p_i(w_2) \) (the shaded element of the probability file for class 2 located at the central position of
the window) and $p_{i,3}(w_4)$ (the shaded element of the probability file for class 4 located at the upper right corner of the window, being the third neighbor of the central pixel) is summed to the element at the position (2, 4, 3) of the compatibility coefficient matrix. A 3 by 3 window is moved over the entire file while the $i$ takes the locations of all the elements until all mutual information is accounted for in an accumulative mode. The other operations, like the summation of the probabilities over the entire file for a particular category $k$ and the logarithmic operation, are quite straightforward without the need of further explanation.

Figure 3-7. Specific neighborhood for the pixels located at the corners and edges of the probability file.

When moving the window over the probability file to summarize the mutual information, we do not have a standard neighborhood for those pixels occurring at the edges and corners. To deal with these situations, the difficulty in finding the standard neighborhood is circumvented as follows: we consider only the subset of the neighborhood related to each specific situation. Taking the upper left corner and the upper edge in the Figure 3-7 as examples, we consider only the neighbor 5, 6, 8, and 9 for the upper corner and neighbor 4, 5, 6, 7, 8, 9 for the upper edge, because only the information
contained in these neighboring pixels has the contribution to the construction of compatibility coefficients. That is, the summation is operated only to those $R_{ij}(w_j, w_k)$'s related to neighbor 5, 6, 8, and 9 for the upper left corner and 4 through 9 for the upper edge while the other $R_{ij}(w_j, w_k)$'s remain unchanged under these particular situations.

3.7 Iterative Refinement of Classification - Probabilistic Relaxation Procedure

Relaxation is one of the best known techniques that have been proposed to exploit contextual information. Conceptually, it adjusts iteratively the initial estimates of the class membership probabilities by reference to the spatial context contained in the neighborhood of a central pixel before the class membership is decided by the maximum likelihood decision rule.

We have defined the R operator in the previous section to what we would like it to behave when taking the contextual information into account of probability adjustment (Figure 3-5). This behavior, however, is only very general in its concept; we need to define it mathematically and have it function such that it really makes appropriate contribution to the class membership probabilities after each iteration. After the definition of compatibility by Equation (3-35), we can proceed to define the R operator mathematically by combining the compatibility coefficient and the current probability file, or more precisely the probability file from the (l-1)th iteration, in an appropriate way.

Let us denote our R operator of the lth iteration for the pixel i as being the class j as $q_i^{(l)}(w_j)$, then,
\[ q_{i}^{(l)}(w_{j}) = \sum_{\alpha=1}^{p} d_{i,\alpha} \left[ \sum_{k=1}^{m} R_{i,\alpha}(w_{j}, w_{k}) p_{i,\alpha}^{(l)}(w_{k}) \right] \]  

(3-36)

where \( d_{i,\alpha} \) is the weighting factor and satisfies \( \sum d_{i,\alpha} = 1 \). Roughly speaking, they function just like the way the a priori probabilities do to the decision of class membership by Equation (3-12). Again, when we have no reason to favor any more particular contribution of a specific neighbor of pixel \( i \) to the operator \( R \) besides that the statistical compatibility has done, we assume that all \( d_{i,\alpha} \)'s are equal, thus we have:

\[ q_{i}^{(l)}(w_{j}) = \sum_{\alpha=1}^{p} \sum_{k=1}^{m} R_{i,\alpha}(w_{j}, w_{k}) p_{i,\alpha}^{(l)}(w_{k}) \]  

(3-37)

By the remarks in the preceding section, the \( q_{i}^{(l)}(w_{j}) \)'s behave just the way we want to the adjustment in \( p_{i}^{(l)}(w_{j}) \) for the \( l \)th iteration. Intuitively, when \( p_{i,\alpha}^{(l)}(w_{j}) \) at \( \alpha \) neighbor of pixel \( i \) is high and \( R_{i,\alpha}(w_{j}, w_{k}) \) is very positive or negative, then the probability of category \( k \) at the \( \alpha \) neighbor will make a substantial positive or negative contribution to \( q_{i}^{(l)}(w_{j}) \); when \( p_{i,\alpha}^{(l)}(w_{j}) \) at the neighbor \( \alpha \) is low, the probability of category \( k \) at the neighbor \( \alpha \) will make little contribution to \( q_{i}^{(l)}(w_{j}) \), irrespective of the value of \( R_{i,\alpha}(w_{j}, w_{k}) \), corresponding the behavior in Figure 3-5.

To make Equation (3-37) more clear, let us spread it out to see how the summation proceeds. After spreading the inner summation,

\[ q_{i}^{(l)}(w_{j}) = \sum_{\alpha=1}^{9} \left[ R_{i,\alpha}(w_{j}, w_{1}) p_{i,\alpha}^{(l)}(w_{1}) + R_{i,\alpha}(w_{j}, w_{2}) p_{i,\alpha}^{(l)}(w_{2}) + \ldots + R_{i,\alpha}(w_{j}, w_{m}) p_{i,\alpha}^{(l)}(w_{m}) \right] \]  

(3-38)

and then the outer summation takes place as:
\[ q_i^{(l)}(w_j) = \]
\[ R_{i,1}(w_j, w_1)p_{i,1}^{(l)}(w_1) + R_{i,1}(w_j, w_2)p_{i,1}^{(l)}(w_2) + \ldots + R_{i,1}(w_j, w_m)p_{i,1}^{(l)}(w_m) + \]
\[ R_{i,2}(w_j, w_1)p_{i,2}^{(l)}(w_1) + R_{i,2}(w_j, w_2)p_{i,2}^{(l)}(w_2) + \ldots + R_{i,2}(w_j, w_m)p_{i,2}^{(l)}(w_m) + \]
\[ \vdots \]
\[ R_{i,g}(w_j, w_1)p_{i,g}^{(l)}(w_1) + R_{i,g}(w_j, w_2)p_{i,g}^{(l)}(w_2) + \ldots + R_{i,g}(w_j, w_m)p_{i,g}^{(l)}(w_m) + \]
\[ \vdots \]
\[ R_{i,9}(w_j, w_1)p_{i,9}^{(l)}(w_1) + R_{i,9}(w_j, w_2)p_{i,9}^{(l)}(w_2) + \ldots + R_{i,9}(w_j, w_m)p_{i,9}^{(l)}(w_m) \]

(3.39)

The program takes summation along the columns instead of rows because it is more straightforward and the algorithm is more easily to be implemented.

As suggested by the functioning of the \( R \) operator, we update probabilities by:

\[ p_i^{(l+1)}(w_j) = p_i^{(l)}(w_j) + q_i^{(l)}(w_j), \text{ where } j = 1, 2, \ldots, m. \]  

(3.40)

However, this definition does not guarantee that the \( p_i \)'s in the \( l \)th iteration remain non-negative because the substantial negative contribution of \( q_i \)'s may occur to a small \( p_i \)'s of the \((l-1)\)th iteration. For a non-negative \( p_i \), we may add \( m \) to \( q_i \)'s as:

\[ p_i^{(l+1)}(w_j) = p_i^{(l)}(w_j) \left[ m + q_i^{(l)}(w_j) \right] \]  

(3.41)

and then apply normalization to \( p_i \)'s for the next iteration:

\[ p_i^{(l+1)}(w_j) = \frac{p_i^{(l)}(w_j) \left[ m + q_i^{(l)}(w_j) \right]}{\sum_{k=1}^{m} p_i^{(l)}(w_k) \left[ m + q_i^{(l)}(w_k) \right]} \]  

(3.42)

The Equation (3.42) is our standard equation for updating the probabilities iteratively.
Here the numerator of Equation (3-42) serves to guarantee that the \( p_j \)'s in the current iteration remain non-negative since readily \( q_j \)'s are in the range \([-m, m]\) provided that \( \sum d_{i,j} = 1 \), so that \( m + q_j^{(l)}(w_j) \) is non-negative; while the denominator serves that the sum of \( p_j \)'s is still one. The discussion on the behavior of operator \( R \) still applies to the updating rule by Equation (3-42): a very positive or negative contribution of the statistical compatibility to \( q_j \)'s introduces an increase or decrease to \( p_j \)'s since the \( p_j^{(l+1)} \)'s are obtained by multiplying \( p_j^{(l)} \) by \( m + q_j^{(l)}(w_j) \), whereas a small contribution to \( q_j \)'s commits only little change to \( p_j \)'s.

The process is the same with the extraction of the mutual information constructing the compatibility coefficients, a 3 by 3 window is moved over the entire probability file until all probabilities for class 1 through \( m \) are updated. Again, the probabilities at the corner and edge positions are updated by considering the contextual information contained in the specific neighborhood only, as shown in Figure 3-7.

The convergence properties of the probabilistic relaxation procedure is rather mathematically oriented and will not be discussed in detail in this thesis. For those interested readers, see Zucker, Krishnamurthy and Haar (1978), and Eklundh and Rosenfeld (1978).

### 3.8 Accuracy Assessment - Error Matrix

The simplest method of evaluating a classified map is to compare it with a reference map in respect to the areas assigned to each category. The result of such comparison is to report the extent of the agreement between two maps with respect to the total area. In this procedure, called "non-site-specific" accuracy assessment, the area of a category is calculated by summing up the numbers of pixels pertaining to that category over the two
maps and then the accuracy is reported based on the area agreement. The most serious drawback of the non-site-specific procedure is that it does not take into account the area change of each category after classification. This drawback is apparent when we inspect the situation in Figure 3-8.

Image A

![Image A](image-a.png)

F=20% W=16% A=64%

Image B

![Image B](image-b.png)

F=20% W=16% A=64%

Figure 3-8. Non-site-specific accuracy assessment.

Figure 3-9. Site-specific accuracy assessment.

The category distribution of image A of Figure 3-8 is apparently different from that of image B; the underestimation of "Forest" in one part of image A can compensate for overestimation of "Forest" in another part of image B, while the area estimation for both image A and image B keeps the same, even though the placement of boundaries differs greatly. The result from this comparison is a "perfect" classification while serious
errors made in the classification are not revealed. This method of comparison could be useful in a large area, small scale agriculture or forest inventory, but it has almost no value as a serious attempt at accuracy assessment for the approach that needs an accurate and efficient method of error assessment to evaluate the accuracy after each refinement of classification, because of its misleading nature in reporting the errors made in the classification.

Table 3-1. Example of an error matrix.

<table>
<thead>
<tr>
<th>IMAGE TO BE EVALUATED</th>
<th>CLASS 1</th>
<th>CLASS 2</th>
<th>CLASS 3</th>
<th>CLASS 4</th>
<th>CLASS 5</th>
<th>CLASS 6</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLASS 1</td>
<td>150</td>
<td>21</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>30</td>
<td>225</td>
</tr>
<tr>
<td>CLASS 2</td>
<td>0</td>
<td>730</td>
<td>93</td>
<td>14</td>
<td>115</td>
<td>21</td>
<td>973</td>
</tr>
<tr>
<td>CLASS 3</td>
<td>33</td>
<td>121</td>
<td>320</td>
<td>23</td>
<td>54</td>
<td>43</td>
<td>594</td>
</tr>
<tr>
<td>CLASS 4</td>
<td>3</td>
<td>18</td>
<td>11</td>
<td>83</td>
<td>8</td>
<td>3</td>
<td>126</td>
</tr>
<tr>
<td>CLASS 5</td>
<td>23</td>
<td>81</td>
<td>12</td>
<td>4</td>
<td>350</td>
<td>13</td>
<td>483</td>
</tr>
<tr>
<td>CLASS 6</td>
<td>39</td>
<td>8</td>
<td>15</td>
<td>3</td>
<td>11</td>
<td>115</td>
<td>191</td>
</tr>
<tr>
<td>TOTALS</td>
<td>248</td>
<td>979</td>
<td>451</td>
<td>134</td>
<td>555</td>
<td>225</td>
<td>1748</td>
</tr>
</tbody>
</table>

As an alternative, the site-specific accuracy assessment is used in this approach (Figure 3-9). In a site-specific procedure, the comparison between classified map and reference map is made on a pixel by pixel basis to accumulate information concerning the correspondence of the two maps at the specific locations. The standard form for reporting site-specific error is the error matrix, sometimes referred as "confusion
matrix" by different authors, because it identifies not only the overall errors for each category, but also the misclassification of a category into another. An error matrix consists of an m+1 by m+1 array, where m is equal to the number of categories on both images (Table 3-1). The left hand side of Table 3-1 is labeled with the categories on the reference image, or "correct" classification; the upper edge is labeled with the same number of categories and by the same order with those listed on the left column of the Table 3-1, referring to those categories on the image to be evaluated.

Inspection of the matrix shows how the matrix elements represent actual classification on the landscape. We see that there are 225 pixels for class 1 in the reference image (the far right element of the first row); of those 225 pixels from class 1, 150 are classified as what they are on the image to be evaluated. Reading succeeding values along the first row, we see next that of those 75 pixels that are classified incorrectly, 21 of them into class 2, 0 into class 3, 7 into class 4, 17 into class 5, and 30 into class 6 on the image to be evaluated. Reading across each row, then we know how a classification scheme assigned those pixels that actually belong to each of the categories as they occur in the landscape.

Inspection of the matrix elements along the column reveals the information about the actual membership of pixels on the reference map when they are actively assigned by a classification scheme incorrectly into another category. For example, reading across the first column, of 248 pixels assigned in classification to class 1, 98 of them belong to categories other than class 1, as 0 to class 2, 33 to class 3, 3 to class 4, 23 to class 5, and 39 to class 6 on the reference image. Reading across each column, then we know how the classification scheme assigned pixels occurring at the landscape to those categories that they actually do not belong to.

Some other information is summarized by the error matrix. The column of totals on the far right gives the total numbers of categories as recorded on the reference image.
The bottom row gives the total numbers of the categories as assigned by the classification scheme on the image to be evaluated. The diagonal elements represent the correctly classified pixels and the total is located at the lower right corner of the matrix as the element \((m+1, m+1)\), thus the percentage correct is derived by dividing element \((m+1, m+1)\) by the sum of the far right column, excluding the element \((m+1, m+1)\).

Table 3-2. Errors of omission and errors of commission.

<table>
<thead>
<tr>
<th></th>
<th>Error of omission</th>
<th>Error of commission</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>75/225 = 33%</td>
<td>98/225 = 44%</td>
<td>150/225 = 67%</td>
</tr>
<tr>
<td>Class 2</td>
<td>243/973 = 25%</td>
<td>249/973 = 26%</td>
<td>730/973 = 75%</td>
</tr>
<tr>
<td>Class 3</td>
<td>274/594 = 46%</td>
<td>131/594 = 22%</td>
<td>320/594 = 54%</td>
</tr>
<tr>
<td>Class 4</td>
<td>43/126 = 34%</td>
<td>51/126 = 41%</td>
<td>83/126 = 66%</td>
</tr>
<tr>
<td>Class 5</td>
<td>133/483 = 28%</td>
<td>205/483 = 42%</td>
<td>350/483 = 73%</td>
</tr>
<tr>
<td>Class 6</td>
<td>76/191 = 40%</td>
<td>110/191 = 58%</td>
<td>115/191 = 60%</td>
</tr>
<tr>
<td>Overall</td>
<td>844/2592 = 33%</td>
<td>844/2592 = 33%</td>
<td>1748/2592 = 67%</td>
</tr>
</tbody>
</table>

The error matrix is essential for any serious study of accuracy in that it reveals two kinds of errors: the error of omission and the error of commission. The error of omission for class \(i\) is the assignment of areas of class \(i\) on the ground to the class \(j\) on the map, in other words, an area of "real" class on the ground is omitted from the map; instead, this area is represented on the map by a number of pixels pertaining to other categories. This error of omission for class \(i\), on the other hand, is defined as the error for commission for class \(j\) (the error rate can be different), as the analyst's error in this instance has been to commit an error actively by assigning a region of class \(j\) to a wrong
category, for example, the class i. The distinction between error of omission and error of commission is essential. In the non-site-specific accuracy assessment, the tabulation of error of matrix reveals that the achievement of 100% correct classification for class i by area is meaningless because there are high errors of commission for class i and high errors of omission for other categories.

By reading across the row of the error matrix, we have the errors of omission for each of the categories; in contrast, reading across the column, then we have the errors of commission for each of categories. Table 3-2 summarizes two kinds of errors by category. For example, 33% of the errors of omission for class 1 gives the information that 33% (75 out of 225) of pixels for the class 1 on the reference image are assigned to other categories on the image to be evaluated; while 44% of the errors of commission means that 44% (98 out of 225) of pixels for class 1 on the classified image actually belong to other categories on the landscape. The overall correct for class 1 is 67%. Note that in non-site-specific accuracy assessment, the accuracy for class 1 is 90.7% by area (225 pixels on the reference image and 248 pixels on the classified image). Here, the misleading nature of the non-site-specific accuracy is apparent.

3.9 Summary

Some of the basic concepts and statistical models related to this study have been introduced in this chapter, while efforts have been tried to make the discussion as non-mathematical as possible by explaining the functioning of these models more literally. The programming is made based on these models to accomplish the goals of this study. For the notational convenience, the mathematical symbols appearing in this chapter are summarized in Table 3-3 by the order of their appearance in the chapter.
Table 3-3. Notational index.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>Number of land cover categories to be identified by a classification scheme.</td>
</tr>
<tr>
<td>n</td>
<td>Number of variables involved in a multi-variable measurement.</td>
</tr>
<tr>
<td>( \mu_i )</td>
<td>Estimated mean value of mono-band measurement for class i.</td>
</tr>
<tr>
<td>q_i</td>
<td>Number of training pixels for class i.</td>
</tr>
<tr>
<td>x_j</td>
<td>Mono-band gray level of the pixel j.</td>
</tr>
<tr>
<td>( \beta_i^2 )</td>
<td>Estimated mono-band variance of measurement for class i.</td>
</tr>
<tr>
<td>( \mu_{ij} )</td>
<td>Estimated mean value on band j for class i.</td>
</tr>
<tr>
<td>x_{jk}</td>
<td>Gray level on band j for pixel k in multi-band measurement.</td>
</tr>
<tr>
<td>( \beta_{ij} )</td>
<td>Measurement covariance between band j and band l for class i.</td>
</tr>
<tr>
<td>U_i</td>
<td>Mean vector for class i.</td>
</tr>
<tr>
<td>( \Sigma_i )</td>
<td>Covariance matrix for class i.</td>
</tr>
<tr>
<td>( p(x</td>
<td>w_i) )</td>
</tr>
<tr>
<td>( \exp[\ ] )</td>
<td>( e ) (the base of natural logarithms) raised to the indicated power in the bracket.</td>
</tr>
<tr>
<td>X</td>
<td>Data vector of the multispectral measurement.</td>
</tr>
<tr>
<td>( p(X</td>
<td>w_i) )</td>
</tr>
<tr>
<td>(</td>
<td>\Sigma_i</td>
</tr>
<tr>
<td>( \Sigma_i^{-1} )</td>
<td>Inverse of matrix ( \Sigma_i ).</td>
</tr>
<tr>
<td>( A^T )</td>
<td>Transpose of matrix A.</td>
</tr>
<tr>
<td>N(( \mu_i, \Sigma_i ))</td>
<td>Normal density function with mean vector ( \mu_i ) and covariance matrix ( \Sigma_i ).</td>
</tr>
<tr>
<td>( p(w_i) )</td>
<td>A priori probability associated with class i.</td>
</tr>
<tr>
<td>( p(w_i</td>
<td>X) )</td>
</tr>
<tr>
<td>( p(X) )</td>
<td>Probability that a measurement vector X is observed.</td>
</tr>
<tr>
<td>D_{ij}</td>
<td>Divergence between class i and class j.</td>
</tr>
<tr>
<td>L(X)</td>
<td>Logarithmic-likelihood ratio of vector X.</td>
</tr>
<tr>
<td>f_i(x_1, x_2, ..., x_K, ..., x_n)</td>
<td>Measurement vector for feature i on n variables.</td>
</tr>
<tr>
<td>tr(A)</td>
<td>Trace of the matrix A.</td>
</tr>
<tr>
<td>D_{ij}(x_k)</td>
<td>Divergence between class i and class j with measurement ( x_k ).</td>
</tr>
<tr>
<td>J_{ij}</td>
<td>J-M distance between class i and class j.</td>
</tr>
<tr>
<td>R</td>
<td>Relaxation operator.</td>
</tr>
<tr>
<td>( p_i(w_k) )</td>
<td>Normalized probability of a given pixel at location i to be class k.</td>
</tr>
</tbody>
</table>
$(i, \partial)$ Subscript for a specific location at the neighbor $\partial$ of the central pixel $i$.

$R_{i, \partial}(w_j, w_k)$ Compatibility coefficient between category $j$ at the location $i$ with category $k$ at the neighbor $\partial$ of the pixel $i$.

$p(w_k)$ Estimation of any pixel as being the category $k$.

$p_{i, \partial}(w_j, w_k)$ Joint probability of a pair of pixels with pixel $i$ belonging to category $W_j$ and the pixel at the neighbor $\partial$ of pixel $i$ belonging to category $W_k$.

$p_{i, \partial}(w_j | w_k)$ Conditional probability that pixel $i$ be category $W_j$ given that the pixel at the neighbor $\partial$ of the pixel $i$ be the category $W_k$.

$I(A)$ Amount of information received as being told that $A$ has occurred.

$I(A, B)$ Mutual information about contribution of $B$ to the information $A$.

$s$ The number of iteration for classification refinement.

$q_{i(l)}(w_j)$ Relaxation operator of the $l$th iteration for the pixel $i$ as being class $j$.

$d_{i, \partial}$ Weighting factor to account for specific contribution of the pixel at the neighbor $\partial$ of the pixel $i$ to the operator $R$.

$p_{i(l)}(w_j)$ Probability of pixel to be class $j$ at the $l$th iteration.
CHAPTER IV TRAINING DATA MANIPULATION

4.1 On the Selection of Training Data

As required by this study, the selection of training data has to be in compliance with the rules listed in the Section 2.2 of Chapter 2 in order to obtain a meaningful classification by an automated approach. The careful selection of training data following the set-up rules serves two major theoretical purposes. First, they should be able to reflect the magnitude of spectral differences of one land cover category from others by means of statistics created from spectral measurements; the compliance of training data with the selection criteria guarantees that the training data give the best separability measurements in terms of statistical distance. Second, they should be the real representation of their corresponding land cover categories in terms of statistical patterns. These requirements are essentially important in obtaining the moderately accurate initial estimate of class membership probabilities of each class, a key requirement for the relaxation procedure.

Ten land cover categories defined and described in the Chapter 2 are selected as the initial classes that could possibly be separated spectrally. These classes can be visually differentiated from both TM imagery and color infrared aerial photo based on their spectral and spatial patterns. In this point, the efforts are mainly emphasized on the rules associated with the selection of training areas.

During the training area selection, the rule 3 can only be approached as closely as possible because it is related with the spectral homogeneity of a training area. A fairly good representation of the measure of homogeneity is the standard deviation listed in Table 4-1. The more homogeneous a training area is, the smaller the standard deviation,
like the instance of WATER; the more heterogeneous a training area is, the greater the standard deviation, like the instance of INDTR.

Table 4-1. The standard deviation extracted from the 6-band training data.

<table>
<thead>
<tr>
<th>BAND</th>
<th>WATER</th>
<th>GOLFC</th>
<th>GRASS</th>
<th>VTREE</th>
<th>BAREG</th>
<th>COMWE</th>
<th>COMCR</th>
<th>INDTR</th>
<th>RDOLD</th>
<th>RDNEW*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.96</td>
<td>12.26</td>
<td>2.82</td>
<td>2.15</td>
<td>11.40</td>
<td>15.12</td>
<td>20.56</td>
<td>28.22</td>
<td>10.57</td>
<td>10.95</td>
</tr>
<tr>
<td>2</td>
<td>3.26</td>
<td>6.80</td>
<td>1.55</td>
<td>1.62</td>
<td>7.46</td>
<td>8.64</td>
<td>11.89</td>
<td>17.84</td>
<td>6.09</td>
<td>6.25</td>
</tr>
<tr>
<td>3</td>
<td>4.93</td>
<td>10.60</td>
<td>2.84</td>
<td>2.32</td>
<td>11.28</td>
<td>12.89</td>
<td>17.05</td>
<td>23.24</td>
<td>9.67</td>
<td>9.90</td>
</tr>
<tr>
<td>4</td>
<td>5.37</td>
<td>18.40</td>
<td>6.01</td>
<td>7.90</td>
<td>7.39</td>
<td>9.59</td>
<td>13.82</td>
<td>13.43</td>
<td>6.36</td>
<td>7.45</td>
</tr>
<tr>
<td>5</td>
<td>7.01</td>
<td>10.42</td>
<td>5.45</td>
<td>6.56</td>
<td>9.23</td>
<td>12.78</td>
<td>19.54</td>
<td>21.22</td>
<td>9.75</td>
<td>10.06</td>
</tr>
<tr>
<td>7</td>
<td>4.79</td>
<td>7.75</td>
<td>3.62</td>
<td>3.60</td>
<td>6.28</td>
<td>8.37</td>
<td>12.24</td>
<td>16.56</td>
<td>6.54</td>
<td>7.62</td>
</tr>
</tbody>
</table>

The rule 5 is satisfied by controlling the size of training area, allowing enough number of pixels to be included in creating the training statistics (Table 4-2). The area with higher homogeneity requires relatively small number of training pixels, while the

Table 4-2. Training pixel number (TPN) for different land cover categories.

<table>
<thead>
<tr>
<th></th>
<th>WATER</th>
<th>GOLFC</th>
<th>GRASS</th>
<th>VTREE</th>
<th>BAREG</th>
<th>COMWE</th>
<th>COMCR</th>
<th>INDTR</th>
<th>RDOLD</th>
<th>RDNEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPN</td>
<td>51</td>
<td>162</td>
<td>179</td>
<td>42</td>
<td>49</td>
<td>189</td>
<td>159</td>
<td>142</td>
<td>225</td>
<td>282</td>
</tr>
</tbody>
</table>

* Note: for the definition of these land cover categories, see Table 2-1 The Initial Definition of the Land Cover Categories of Section 2.2 in page 27 for details.
Figure 4-1. Histograms of gray level distribution from WATER training data set.

relatively heterogeneous areas should contain more pixels because all spectral components in the area need to be accounted for.

The rule 4 is generally satisfied of all training sets as illustrated by the histograms from two typical training data sets, WATER and INDTR, shown in Figures 4-1 and 4-2. In Figure 4-2, the normal distribution of gray level is not so typical like that of Figure 4-1. It should be pointed out, however, that this slight violation over few bands contributes no major influence on the behavior of a maximum likelihood classification procedure.
4.2 Matching Spectral Classes with Informational Classes

Conceptually speaking, the selection of initial land cover categories reflects only the intention that the classification by an automated approach could possibly reach such detail. Whether such detail can be reached or not, however, depends to what degree these land cover categories are confused from each other in an n-dimensional classification space. When the confusion reaches to certain degree, it is impossible to achieve a
meaningful classification regardless of how efficient and accurate a classification procedure is. Up to this point, we have treated the informational classes and spectral classes as one to one correspondence. What we need to do is to identify those spectral classes that are spectrally separable based on training statistics and then match these spectral classes with their corresponding informational classes. This is a reliable measure of how much detail could be possibly reached by an automated procedure that put much consideration on the spectral reflectance.

Table 4-3. J-M distance between different land cover categories.

<table>
<thead>
<tr>
<th></th>
<th>WATER</th>
<th>GOLFC</th>
<th>GRASS</th>
<th>VTREE</th>
<th>BAREG</th>
<th>COMWE</th>
<th>COMCR</th>
<th>NDTR</th>
<th>RDO</th>
<th>RDNE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WATER</td>
<td>0</td>
<td>733</td>
<td>1175</td>
<td>1012</td>
<td>1143</td>
<td>1605</td>
<td>1420</td>
<td>1556</td>
<td>759</td>
<td>788</td>
</tr>
<tr>
<td>GOLFC</td>
<td>0</td>
<td>105</td>
<td>470</td>
<td>845</td>
<td>362</td>
<td>353</td>
<td>430</td>
<td>206</td>
<td>201</td>
<td></td>
</tr>
<tr>
<td>GRASS</td>
<td>0</td>
<td>409</td>
<td>1025</td>
<td>496</td>
<td>133</td>
<td>133</td>
<td>555</td>
<td>504</td>
<td>652</td>
<td>251</td>
</tr>
<tr>
<td>VTREE</td>
<td>0</td>
<td>1133</td>
<td>555</td>
<td>504</td>
<td>652</td>
<td>251</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAREG</td>
<td>0</td>
<td>581</td>
<td>416</td>
<td>363</td>
<td>872</td>
<td>767</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMWE</td>
<td>0</td>
<td>79</td>
<td>93</td>
<td>309</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMCR</td>
<td>0</td>
<td>77</td>
<td>382</td>
<td>357</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDTR</td>
<td>0</td>
<td>381</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDO</td>
<td>0</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDNE</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 6-band training data are used as input to the programs based on Equation (3-21) and Equations (3-27) and (3-28) to calculate statistical distances between different land cover categories. The 6-band data set can incorporate as much spectral information as possible to give better differentiation between categories than those by using any band combinations. The results are tabulated in Tables 4-3 and 4-4. Before the actual classification is done, some information about classification result has been
extracted from these tables. For the convenience of discussion, let us check the J-M distance in Table 4-3 by row, and by column when necessary, to see what information it is able to reveal.

Table 4-4. Divergence between different land cover categories.

<table>
<thead>
<tr>
<th></th>
<th>WATER</th>
<th>GOLFC</th>
<th>GRASS</th>
<th>VTREE</th>
<th>BAREG</th>
<th>COMWE</th>
<th>COMCR</th>
<th>INDTR</th>
<th>RDOLD</th>
<th>RDNEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>WATER</td>
<td>0</td>
<td>9604</td>
<td>6430</td>
<td>8185</td>
<td>2993</td>
<td>3290</td>
<td>3870</td>
<td>4350</td>
<td>4520</td>
<td>4109</td>
</tr>
<tr>
<td>GOLFC</td>
<td>0</td>
<td>71</td>
<td>407</td>
<td>2507</td>
<td>735</td>
<td>611</td>
<td>785</td>
<td>295</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>GRASS</td>
<td>0</td>
<td>425</td>
<td>1334</td>
<td>481</td>
<td>560</td>
<td>660</td>
<td>294</td>
<td>252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VTREE</td>
<td>0</td>
<td>693</td>
<td>653</td>
<td>435</td>
<td>1141</td>
<td>423</td>
<td>402</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAREG</td>
<td>0</td>
<td>1245</td>
<td>1228</td>
<td>1073</td>
<td>878</td>
<td>950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMWE</td>
<td>0</td>
<td>6</td>
<td>65</td>
<td>168</td>
<td>352</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMCR</td>
<td>0</td>
<td>53</td>
<td>414</td>
<td>586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INDTR</td>
<td>0</td>
<td>395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDOLD</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDNEW</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the first row, the distance numbers between WATER and any other categories are great enough, that is, in an n-dimensional classification space, WATER is "far" enough from any other classes in terms of statistical distances. This fact is easily understood by the phenomenon that water is a homogeneous body of ground cover bearing unique spectral characteristics. During the actual classification, we expect no or very few misclassifications of water into other categories or vice versa.

In the second and third rows, the distances still keep moderate to significant large, but some anomalies and interesting features should be noted. Firstly, in the row GOLFC and column GRASS, the distance decreases rapidly, reflecting the fact that although the golf course and regular grass show some difference in spatial pattern and maturity in its
development, this difference does not bring any spectral significance characterizing each of these land cover categories; they should then be grouped into the same spectral classes and the corresponding informational classes. Secondly, the distances between GOLFC and GRASS and VTREE still keep large to have grass and tree differentiated. Thirdly, the distance between grass and residential areas, either old or new, is relatively small; this feature is attributed to the spectral characteristics of residential areas. In fact, the reflectance from residential areas is only the mixture from the grass and houses. When grass occupies a large portion of a single pixel, the grass component of spectral reflectance will play a major role in the spectral representation of that particular pixel; when the situation is reversed, the spectral behavior will be more similar with that of houses. This problem is typically represented by the overlap area 1 of Figure 1-3b and we will leave its solution to the relaxation procedure.

The distance numbers in the fourth row still keep the same trend with that of the third row except those at the last two columns. The occurrence of this anomaly can be attributed to the same reason discussed in the last paragraph. We may expect that more classification errors occur between these land cover categories, that is, between the vegetation and residential areas.

In the fifth row, what the distance numbers reveals is that the BAREG can be established as an individual category without risking too much misclassification, because the statistical distances between BAREG and any other land cover categories indicate that although the BAREG does not have such a unique spectral characteristic like that of WATER, it is still located at a specific position in an n-dimensional classification space, fairly "far" from other categories.

Beginning from the sixth row, the distance numbers vary between different land cover categories. The statistical distances between COMWE, COMCR and INDTR are considerably smaller than those between any others. This fact states that they belong to
the same spectral classes in terms of spectral characteristics, though they should be
classified at least as two different land use categories by functions (commercial and
industrial). Unfortunately, we can only group them under the same category according
to their spectral behavior. Up to this point, we have not yet been able to identify the
functioning of a particular land use category as one of the parameters of our
classification procedure, because what we consider as the most important information in
our classification procedure is the roof type of buildings rather than the activities
happening under a particular roof. The distance number in the last row also states the
simple fact that we do not consider RDOLD and RDNEW as separate classes, regardless of
how much spatial pattern can be visually depicted from the imagery when we have no
efficient means to represent quantitatively the spatial patterns. The similar patterns
was found in the divergence.

Table 4-5. J-M distance from different land cover categories after adjusting.

<table>
<thead>
<tr>
<th></th>
<th>COMIN</th>
<th>WATER</th>
<th>GRASS</th>
<th>TREES</th>
<th>BAREG</th>
<th>RESID</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMIN</td>
<td>0</td>
<td>1605</td>
<td>362</td>
<td>555</td>
<td>581</td>
<td>373</td>
</tr>
<tr>
<td>WATER</td>
<td>0</td>
<td>733</td>
<td>1012</td>
<td>1143</td>
<td>788</td>
<td></td>
</tr>
<tr>
<td>GRASS</td>
<td>0</td>
<td>470</td>
<td>845</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREES</td>
<td>0</td>
<td>1133</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAREG</td>
<td>0</td>
<td>767</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESID</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After checking the divergence and J-M distance, the initial ten land cover categories
are regrouped into two major groups based on the spectral behavior: those that are "far"
from each other in an n-dimensional classification space are kept separately as the
individual categories; while those that are too "close" from each other are combined into
the same categories. The results from this comparison are more clearly shown in Tables 4-5 and 4-6. The statistical distances from these tables can be considered as the initial qualitative estimates of the occurrence of misclassification between different land cover types: the greater the distance magnitude is, the higher the risk of misclassification is.

Table 4-6. Divergence from different land cover categories after adjusting.

<table>
<thead>
<tr>
<th></th>
<th>COMIN</th>
<th>WATER</th>
<th>GRASS</th>
<th>TREES</th>
<th>BAREG</th>
<th>RESID</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMIN</td>
<td>0</td>
<td>3290</td>
<td>735</td>
<td>653</td>
<td>1245</td>
<td>352</td>
</tr>
<tr>
<td>WATER</td>
<td>0</td>
<td>9604</td>
<td>8185</td>
<td>2993</td>
<td>4109</td>
<td></td>
</tr>
<tr>
<td>GRASS</td>
<td>0</td>
<td>407</td>
<td>2507</td>
<td>402</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TREES</td>
<td>0</td>
<td>693</td>
<td>950</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAREG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESID</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The definition of the new land cover categories is straightforward from their class names listed in the Tables 4-5 and 4-6.

Note that it is not the attempt of this comparison to establish a threshold to regroup categories based on the statistical distance; this is only the effort to reduce the chance of misclassification as much as possible using a numerical analysis technique. The statistical distances are compared with each other by their relative magnitudes instead of comparing them with certain threshold and the expectation of classification error is also made based on the quantitative representation of spectral similarity between different land cover categories. From this point of view, the statistical distance can really be used as an efficient way to describe quantitatively the spectral behavior of land covers.
4.3 Summary

The training areas are depicted visually from both TM imagery and color infrared aerial photo based on the selection rules listed in Chapter 2. The more closely does the selection of training area follow these rules, the more reliable the training statistics is, which has significant influence over the accuracy of a supervised classification scheme.

The statistics are extracted from 6-band training data to create statistical distance matrices that are used as the basic measurement to describe the spectral differences between different land cover categories. The initial ten land cover categories are regrouped into six individual classes according to their relative position of one from each other represented by statistical distances in a 6-dimensional classification space. The rationale of this comparison is that the spectral similarity can be represented numerically by statistical models and then be easily described without involving too much human opinion resulting from the visual interpretation of the imagery. When we classify land cover categories using an automated approach with much consideration in spectral component, we are expecting a much better result by classifying those categories that have the least spectral similarity.

The major purpose of spectral separability checking is to identify those land cover categories that have much similar spectral patterns; when the spectral confusion reaches such degree like that represented by overlap area 2 of Figure 1-3b, it is hardly possible to identify these categories separately from each other with acceptable error rates by an automated classification procedure to be used in this study. The analysis of divergence and J-M distance provides a much clearer picture on which type of land cover categories can be established as individual ones and which ones can not be; the chance of misclassification a particular land cover category might have; and if misclassification does occur, where it is most likely to occur.
CHAPTER V ANALYSIS

5.1 Probabilistic Relaxation

After solving the spectral signature confusion by the procedure described in the Chapter IV, we now turn to the solution of class membership confusion, which is also introduced in some sense by the spectral confusion but in a manageable degree. During the GML classification, when the class membership is determined by the maximum likelihood decision rule, it does not necessarily mean that this decision is definitely right. From the statistical point of view and the structure of this classifier, those pixels whose class membership have been decided at a particular classification stage still have some possibility to be other classes, while the magnitude of this possibility depends on the amount of contextual information contained in the neighborhood of this particular pixel.

The probabilistic relaxation procedure starts refining the classification based on that done by a GML classifier. The theoretical basis for any improvement of accuracy by this procedure is that the GML itself achieves a reasonable accuracy, otherwise, the relaxation procedure can not be used to achieve a significant accuracy improvement. Unfortunately, there is no absolute standard in the related literature for an acceptable level of classification accuracy. Rather, it depends on the nature of the task and the data and algorithms used. As a practical issue, however, most literature views 70% as the acceptable level of accuracy of a machine-based classification of satellite data. When the classification accuracy is lower than this level, we say that we still have the problem of spectral signature confusion and can not use the relaxation procedure, otherwise, the accuracy is degraded instead of upgraded.
The TM data covering the area shown in Figure 2-1 is classified into six different land cover categories using a GML classification procedure. Two files are created. One is the probability file that is going to be used to calculate statistical compatibility and for further iterations to improve the accuracy. Another file is the classification map created from probability file using maximum likelihood decision rule.

Table 5-1. Error matrix from the GML classification.

<table>
<thead>
<tr>
<th></th>
<th>COMIN</th>
<th>WATER</th>
<th>GRASS</th>
<th>TREES</th>
<th>BAREG</th>
<th>RESID</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMIN</td>
<td>18375</td>
<td>54</td>
<td>1877</td>
<td>39</td>
<td>233</td>
<td>4648</td>
<td>25226</td>
</tr>
<tr>
<td>WATER</td>
<td>15</td>
<td>83</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>107</td>
</tr>
<tr>
<td>GRASS</td>
<td>502</td>
<td>5</td>
<td>6062</td>
<td>275</td>
<td>104</td>
<td>1134</td>
<td>8082</td>
</tr>
<tr>
<td>TREES</td>
<td>40</td>
<td>0</td>
<td>466</td>
<td>2152</td>
<td>0</td>
<td>501</td>
<td>3159</td>
</tr>
<tr>
<td>BAREG</td>
<td>252</td>
<td>0</td>
<td>79</td>
<td>11</td>
<td>625</td>
<td>52</td>
<td>1019</td>
</tr>
<tr>
<td>RESID</td>
<td>1258</td>
<td>0</td>
<td>2609</td>
<td>894</td>
<td>12</td>
<td>23170</td>
<td>27943</td>
</tr>
<tr>
<td>TOTALS</td>
<td>20442</td>
<td>142</td>
<td>11096</td>
<td>3371</td>
<td>974</td>
<td>29511</td>
<td>50467</td>
</tr>
</tbody>
</table>

**ERROR OF OMISSION**

- COMIN 6851/25226 = 27.16%
- WATER 24/107 = 22.43%
- GRASS 2020/8082 = 24.99%
- TREES 1007/3159 = 31.88%
- BAREG 394/1019 = 38.67%
- RESID 4773/27943 = 17.08%

**ERROR OF COMMISSION**

- 2067/25226 = 8.19%
- 59/107 = 55.14%
- 5034/8082 = 62.29%
- 1219/3159 = 38.59%
- 349/1019 = 34.25%
- 6341/27943 = 22.69%

**CORRECT**

- 18375/25226 = 72.84%
- 83/107 = 77.57%
- 6062/8082 = 75.01%
- 2152/3159 = 68.12%
- 625/1019 = 61.33%
- 23170/27943 = 82.92%

OVERALL 15069/65536 = 22.99%

The map is compared pixel by pixel with the manually created ground truth map and an error matrix is compiled to indicate the accuracy of the current classification. This error matrix, or confusion table, after the GML classification is shown in Table 5-1. The overall accuracy of 77.01% is good enough to start the relaxation procedure, though
there are some lower percentages of correct for VTREE and BAREG. This anomaly has no major influence on the overall classification accuracy because only 15.77% of pixels from the entire area under study belong to these land cover categories.

The compatibility coefficient matrix is created from the probability file by Equation (3-35). The elements of this matrix represent the statistical compatibility between different classes in terms of class membership occurrence at a specific neighborhood of a central pixel. These compatibility coefficients are fixed throughout the following iterations. The reason to require fixed coefficients for the entire relaxation procedure is that the probabilities from the previous iteration reflect only the transitional status of class membership during the continuous adjustment of probabilities; if the compatibility coefficients are recalculated after each iteration, they do not represent more precisely the compatibility status than that represented by the initial probabilities from the original GML classification. As pointed out in the Chapter I, this initial probability contains the most information and be able to give more accurate description about the class compatibility. The actual compatibility coefficient matrix is listed in Table 5-2.

The compatibility coefficients and probability file from the previous iteration are used in the current iteration to adjust probabilities and create probability file for the next iteration by Equations (3-36) through (3-42). After each iteration, the probability file is converted by maximum likelihood decision rule to the classification map; this map is then compared each time with the ground truth map to check classification accuracy from the current iteration. An error matrix is compiled for each iteration to compare the improvement of classification accuracy on each land cover category. When there is no more increase in accuracy or the accuracy begins to degrade after current iteration, the relaxation process stops and the classification map from the last probability file is considered as the best result.
Table 5-2. Compatibility coefficients from a 3 by 3 neighborhood.

<table>
<thead>
<tr>
<th>$R_{ij}(1,1)$</th>
<th>$R_{ij}(2,2)$</th>
<th>$R_{ij}(3,3)$</th>
<th>$R_{ij}(4,4)$</th>
<th>$R_{ij}(5,5)$</th>
<th>$R_{ij}(6,6)$</th>
<th>$R_{ij}(7,7)$</th>
<th>$R_{ij}(8,8)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.007967</td>
<td>0.010766</td>
<td>0.007678</td>
<td>-0.070869</td>
<td>-0.003300</td>
<td>-0.001960</td>
<td>0.005191</td>
<td>0.010730</td>
</tr>
<tr>
<td>0.027235</td>
<td>0.032009</td>
<td>0.027981</td>
<td>-0.020748</td>
<td>-0.000300</td>
<td>-0.002014</td>
<td>-0.002014</td>
<td>0.003232</td>
</tr>
<tr>
<td>-0.005163</td>
<td>-0.003920</td>
<td>-0.004984</td>
<td>0.005925</td>
<td>-0.000001</td>
<td>0.005425</td>
<td>0.005425</td>
<td>-0.000172</td>
</tr>
<tr>
<td>-0.049122</td>
<td>-0.056129</td>
<td>-0.054604</td>
<td>0.055850</td>
<td>-0.016584</td>
<td>0.015705</td>
<td>0.015971</td>
<td>-0.000187</td>
</tr>
<tr>
<td>0.051734</td>
<td>0.054825</td>
<td>0.054779</td>
<td>-0.057059</td>
<td>-0.001685</td>
<td>-0.001833</td>
<td>0.003020</td>
<td>0.001076</td>
</tr>
<tr>
<td>-0.003300</td>
<td>0.051275</td>
<td>0.050098</td>
<td>0.000371</td>
<td>0.001875</td>
<td>-0.002089</td>
<td>0.002089</td>
<td>-0.001872</td>
</tr>
<tr>
<td>-0.015684</td>
<td>0.000456</td>
<td>0.002187</td>
<td>-0.001993</td>
<td>0.000476</td>
<td>-0.004256</td>
<td>0.004256</td>
<td>-0.004256</td>
</tr>
<tr>
<td>-0.000560</td>
<td>0.016595</td>
<td>0.016595</td>
<td>-0.002187</td>
<td>0.000476</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.003071</td>
<td>0.003071</td>
<td>0.000476</td>
<td>-0.002187</td>
<td>-0.002187</td>
<td>0.002187</td>
<td>0.002187</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.001833</td>
<td>-0.001833</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.002014</td>
<td>-0.002014</td>
<td>-0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.004256</td>
<td>0.004256</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.005425</td>
<td>0.005425</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.001833</td>
<td>-0.001833</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.002014</td>
<td>-0.002014</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.005425</td>
<td>0.005425</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.001833</td>
<td>-0.001833</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.002014</td>
<td>-0.002014</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.005425</td>
<td>0.005425</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.001833</td>
<td>-0.001833</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>-0.002014</td>
<td>-0.002014</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
<tr>
<td>-0.000540</td>
<td>0.005425</td>
<td>0.005425</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
<td>0.000476</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$R_{ij}(9,9)$</th>
<th>$R_{ij}(8,8)$</th>
<th>$R_{ij}(7,7)$</th>
<th>$R_{ij}(6,6)$</th>
<th>$R_{ij}(5,5)$</th>
<th>$R_{ij}(4,4)$</th>
<th>$R_{ij}(3,3)$</th>
<th>$R_{ij}(2,2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.007967</td>
<td>0.027235</td>
<td>0.027981</td>
<td>0.027981</td>
<td>0.027981</td>
<td>0.027981</td>
<td>0.027981</td>
<td>0.027981</td>
</tr>
</tbody>
</table>

90
During the actual operation, the classification accuracy begins to increase quickly after the first few several iterations and then was slowed down with further adjustment of probabilities. The degree of improvement differs from different categories, depending on their spectral behaviors and the relationship between them. The overall increase of classification accuracy is shown graphically in Figure 5-1 that gives us a general idea on the improvement of classification accuracy after each iteration. After 13 iterations, the overall classification accuracy has reached the maximum up to 88.48% correct. The overall accuracy begins to decrease, however, with the further iterations, though the accuracies from few single categories still keep increasing in the next iteration, which is compensated by the decrease of accuracy from other categories, resulting in an
Table 5-3. Error matrix from the last iteration.

<table>
<thead>
<tr>
<th></th>
<th>COMIN</th>
<th>WATER</th>
<th>GRASS</th>
<th>TREES</th>
<th>BAREG</th>
<th>RESID</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMIN</td>
<td>21886</td>
<td>11</td>
<td>1051</td>
<td>17</td>
<td>62</td>
<td>2199</td>
<td>25226</td>
</tr>
<tr>
<td>WATER</td>
<td>5</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>107</td>
</tr>
<tr>
<td>GRASS</td>
<td>184</td>
<td>1</td>
<td>7077</td>
<td>109</td>
<td>33</td>
<td>678</td>
<td>8082</td>
</tr>
<tr>
<td>TREES</td>
<td>32</td>
<td>22</td>
<td>86</td>
<td>2980</td>
<td>8</td>
<td>31</td>
<td>3159</td>
</tr>
<tr>
<td>BAREG</td>
<td>152</td>
<td>3</td>
<td>31</td>
<td>2</td>
<td>814</td>
<td>17</td>
<td>1019</td>
</tr>
<tr>
<td>RESID</td>
<td>939</td>
<td>3</td>
<td>1443</td>
<td>415</td>
<td>11</td>
<td>25132</td>
<td>27943</td>
</tr>
<tr>
<td>TOTALS</td>
<td>23198</td>
<td>140</td>
<td>9690</td>
<td>3523</td>
<td>928</td>
<td>28057</td>
<td>57989</td>
</tr>
</tbody>
</table>

**ERROR OF OMISSION**

- COMIN: \(\frac{3340}{25226} = 13.24\%\)
- WATER: \(\frac{7}{107} = 6.54\%\)
- GRASS: \(\frac{1005}{8082} = 12.44\%\)
- TREES: \(\frac{179}{3159} = 5.67\%\)
- BAREG: \(\frac{205}{1019} = 20.12\%\)
- RESID: \(\frac{2811}{27943} = 10.06\%\)
- OVERALL: \(\frac{7547}{65536} = 11.52\%\)

**ERROR OF COMMISSION**

- COMIN: \(\frac{1312}{25226} = 5.20\%\)
- WATER: \(\frac{40}{107} = 33.38\%\)
- GRASS: \(\frac{2611}{8082} = 32.33\%\)
- TREES: \(\frac{543}{3159} = 17.19\%\)
- BAREG: \(\frac{114}{1019} = 11.19\%\)
- RESID: \(\frac{2925}{27943} = 10.47\%\)
- OVERALL: \(\frac{7547}{65536} = 11.52\%\)

**CORRECT**

- COMIN: \(\frac{21886}{25226} = 86.76\%\)
- WATER: \(\frac{100}{107} = 93.46\%\)
- GRASS: \(\frac{7077}{8082} = 87.56\%\)
- TREES: \(\frac{2980}{3159} = 94.33\%\)
- BAREG: \(\frac{814}{1019} = 79.88\%\)
- RESID: \(\frac{25132}{27943} = 89.94\%\)
- OVERALL: \(\frac{57989}{65536} = 88.48\%\)

Overall degrading of classification performance. The error matrix from the last iteration is given in Table 5-3.

### 5.2 Error Patterns

With a classification over a 256 by 256 image, it is difficult to discuss the problem of error patterns of a 6-class classification over the entire image without an efficient display device on VAX. Therefore, the discussion of error patterns is focused specifically
on the category with a smaller number of pixels and a predictable spectral behavior. The discussion of error patterns and class membership confusion for other categories is only attempted as more general.

Category WATER, a homogeneous land cover category among others, consists of 107 pixels and it has a greater statistical distance from any other categories in a 6-dimensional classification space (see Tables 4-5 and 4-6). We focus our discussion of error pattern on WATER to trace the incorrectly classified pixels and its possible reason. The experience of operating a probabilistic relaxation procedure on both artificial and real imagery tells us that the more homogeneous a category is in terms of physical appearance and the more predictable its spectral behavior is, the more likely
does the error pattern occur similarly after each iteration. Or, in other words, the improvement of classification accuracy proceeds with the same pattern during each iteration.

In Figure 5-2a is the WATER coverage after the GML classification, representing a 77.57% correct classification. Most incorrectly classified pixels are located at the boundary of the water body. The inspection of the original TM imagery and the aerial photo reveals that some buildings, probably the recreation facilities, are located on the south side of the lake. To the east side of the lower (south) part of the lake are few pieces of grass and few houses surrounded by the grass, giving a typical reflectance of a residential area, though these houses may not function as residential units. Within the extent of water coverage, there are few pixels classified as other categories that were mistaken by the strict decision rule when the GML classification was done. Most probably, the reflectance from the boundary pixels are influenced by the surrounding environment, depending on how much percentage coverage of that particular pixel by either WATER or other categories. In this case, we say that they have a potential to become WATER pixels, provided that proper adjustment is done by taking into account the contextual information. The same situation applies to those pixels surrounded only by WATER pixels. During the iteration, the probabilities of these boundary pixels are updated continuously by the relaxation procedure. The accuracy is increased steadily at the very beginning and then slowed down until it reaches a fixed point before declining. As a result, the classification accuracy is increased from 77.57% to 93.46% for WATER, while those pixels classified by GML as WATER never changed their class membership. In other words, the better result is achieved only by adjusting probabilities of the incorrectly classified pixels instead of the compensation between correctly and incorrectly classified pixels. The remaining incorrectly classified pixels are from the
limitations of either ground truth reference map from manual interpretation or the iteration procedure itself.

The reason for the higher rate of errors of commission for the WATER (see Tables 5-1 and 5-3) is more difficult to explain than that of errors of omission. From the inspection of error matrix, the WATER is more easily to be committed as error pixels to COMIN (commercial/industrial) and vegetation pixels (TREES and GRASS). When they are considered as errors of omission for other categories, however, it is only a minor issue without any importance with regards to the overall accuracy.

Figure 5-3. The reduction of error rate by iterative adjustment of probabilities.

Figure 5-3 gives a more general representation of the error rate reduction in terms of errors of omission through different stages of the relaxation procedure. We focus our discussion on errors of omission, indicating the roles of the errors of commission when necessary, to provide a clearer picture of the relationship between different land cover
categories in terms of their spectral behaviors, their physical distribution on the ground, and their possible influences on the classification accuracy.

From the original GML classification, the error rate of COMIN is 27.16%, while the most errors are committed by RESID (67.84% of error pixels) and GRASS (27.40% of error pixels). After the 13th iteration, the error rate is reduced more than 50% down to 13.24%, with both RESID and GRASS as the most frequently occurring error pixels. Considering the higher frequency of error occurrence at the boundary pixels, the higher percentage of coverage by area by COMIN and RESID (81.13% combined of the entire image) and the highly diversified distribution of grey levels from histogram of COMIN (see Figure 4-2), this error pattern is predictable. In fact, before the relaxation procedure is started, the identity of boundary pixels between GRASS and COMIN has been in a ambiguous status. During the iteration, the contextual information contained in the neighborhood of these pixels is used to increase, or decrease their probabilities of being the separate classes until it reaches a saturated status at the final stage of iteration.

The clarification of identity ambiguity depends on such factors as the dimension of the ambiguous pixels, their actual spectral separability, the accuracy of manual classification, and so forth. For example, large pieces of grass in a commercial area with the dimension of corresponding 10 pixels might not be picked up as a separate category in the reference map because of the limitation of manual classification. During the iteration, however, the 3 by 3 neighborhood is used to accumulate the influence of contextual information on the central pixel, and in this case, the area of 10 pixel is large enough to be a neighborhood that have much influence on the central pixel to change its identity toward the direction opposite to the background of that particular neighborhood, thus creating error pixels when compared with the reference map.

Another example of this phenomenon is that a small commercial building with its vegetative surrounding is spectrally similar to a residential unit. The error is thus
easily to be committed by RESID to COMIN, either within the COMIN or between the boundaries of two land cover categories, among both the former is the common situation within the study area. Generally speaking, the error patterns from these examples are major ones for each land cover category. Comparing with the boundary pixels, the error pixels occupying certain dimension and surrounded by a major cover category are not easily to be corrected by relaxation procedure because these pixels sometimes may not really be the "error" pixels in nature by reflectance; in the practice of visual interpretation we prefer to include them into their background land use category by the actual function of these pixels using a subjective land classification system. This problem of dual identity can only be approached by the system with higher intelligence.

The category GRASS starts at 24.99% error rate and ends with the rate down to 12.44%. Most confusion has occurred between GRASS and COMIN (24.85% of the original error pixels) and between GRASS and RESID (56.14% of the original error pixels). With regarding to the spectral separability checking, the confusion should not occur between GRASS and COMIN because of higher degree of spectral separability between them (Tables 4-5 and 4-6). The reasonable reason for this occurrence is that most error pixels are boundary pixels, considering the dimension of COMIN, and they can be reduced by iterations. This assumption is supported by the fact that after 13 iterations, the composition of error pixels has changed to 18.31% between the GRASS and COMIN, and 67.46% between the GRASS and RESID in terms of errors of omission. The later higher percentage reflects the fact that the similarity of the spectral behavior between land cover categories GRASS and RESID makes it more difficult for RESID to withdraw from GRASS territory as that COMIN does to GRASS during iteration.

TREES has presented an interesting error pattern different from any other categories: it starts at the higher error rate (31.88%) and ends with the lowest rate (5.67%) among others. This example gives us a chance to look more closely into the
function the relaxation procedure performs. In the area under study, it is unlikely that trees can develop into forest except those in the public parks and along the creeks. Besides these forests, there are some places that are covered by relatively large areas of trees, though they are not in the scale of forest. Another pattern is that trees are scattered within each type of land cover category, we probably do not want to classify them as separate category other than its background category because they occupy only small areas. During the GML classification, however, these individual groups of trees are classified into the same category with that of forest, which has created great difference from the manually produced reference map. Then in the iteration, these individual groups of trees with less extent of coverage, say, in the dimension of a couple of pixels, are merged into their background categories based on the contextual information summarized in the form of compatibility coefficients. In this case, if a group of pixels occupy a substantially large area, or the individual pixels have a very strong spectral characteristics of tree within the IFOV, they remain unchanged, otherwise, the identity of individual pixel is questioned by the information contained in its neighborhood which is definitely another category without ambiguity. This process is continued until those pixels with less dimension of coverage are merged into their background categories. Only forests and few individual trees are left untouched, resulting in an error rate as low as 5.67%.

BAREG is a category with much complexity. Its original error rate is as high as 38.67% and is ended as the highest error rate: 20.12%. As an individual category, its physical nature makes it difficult to be defined and its spectral behavior under the different physical conditions makes it difficult to be differentiated on certain bands. For example, when a piece of bare ground is very wet, it is more likely to behave like vegetation on band 5 and band 7 of TM; when it is very dry, its behavior is more close to that of commercial building on bands 1, 2, and 3. Furthermore, the primitive
development of vegetation on bare ground makes it more likely to have a transitional spectral behavior. These factors make the BAREG have different associations with different land cover categories according to its physical dimension and status. With the combination of these difficulties in mind, it is not a big surprise that the reduction of error rate is not so impressive after the first few iterations. Only the pixels that present the similar spectral behavior with those training pixels remain unchanged and their boundary pixels have the chance to be adjusted into the category BAREG. Others have fell into different categories, depending on different physical situation and the contextual information available.

The RESID gives the lowest original error rate among others while COMIN and GRASS pixels are the major committed pixels. With the reasoning of error patterns in the previous paragraphs, it is not difficult to understand some reasons that have created such error pattern. First, the confusion between COMIN and RESID is created by the physical appearance. For example, the buildings occupying certain dimension in the residential areas, like children's playing ground, churches, smaller shopping facilities, and schools might not be counted as separate COMIN category by reference map because of the small dimension. But, when they do present different spectral pattern from their background and occupy a relatively large dimension, the procedure itself is not able to adjust the probabilities so much that most of them be merged into the background categories. As the result, the error rate of COMIN to RESID drops only 25%, from 1258 pixels to 939 pixels in terms of errors of omission. Second, the actual spectral behavior of RESID is the mixture of vegetation, mainly grass, and roof types, and the spectral behavior of a particular pixel is determined by the composition of land coverage within the IFOV. Again, when the dimension of grass is substantially small, it is merged into the corresponding background category. The class membership of those pixels at the boundaries of these two land cover categories, on the other hand, can also be adjusted like
what has been done to other categories. By this pattern, the number of error pixels has dropped from the original 2609 to 1443 after the 13th iteration.

5.3 Summary

The GML classification is done with a 6-band TM data using a 6-class training data set. The classified map from the GML is then compared with the reference map on the pixel by pixel basis to check the accuracy and 77.01% overall accuracy has been achieved up to this point. After the classification, the original six class probability file is retained, from which the compatibility coefficient matrix is created that is then used as the summary of the contextual information for the probabilistic relaxation procedure.

A nearest neighborhood procedure (8 neighbors) is used to create compatibility coefficients and to update probabilities because it gives the best result by iteration among other selections of neighborhood. After each iteration, the updated probability file is converted into classification map by the maximum likelihood decision rule and then compared repeatedly with a reference map to improve the classification after each iteration. The compatibility coefficient matrix is fixed throughout the iteration because it contains the most information on compatibility summarized from the original probability file. After thirteen iterations, the 88.48% overall accuracy has been achieved, representing a 14% improvement of the overall performance.

The category WATER is selected as an example to illustrate the error patterns. Most errors are found to occur at the boundaries between different land cover categories and in less degree do they occur within the major categories. The reason for this error pattern is that the GML classifier can not solve the problem of identity ambiguity of those pixels at the boundaries and scattered within the major category. Some of these
error pixels are reclassified into corresponding correct categories by adjusting their probabilities while some of them remain untouched as the classification errors.

Two changing patterns are analyzed on the category by category basis. The degree of success of the relaxation procedure depends on such factors as the physical appearance and the spectral behavior of different land cover categories, the quality of the original training data, the accuracy of the GML classification and the reference map, and the limitation of the probabilistic relaxation procedure itself.
6.1 On the Classification Scheme

Multispectral classification has played an important role in the application of remote sensing technology to geographical research. In the last two decades, many techniques have been developed by researchers and scientists from different disciplines to solve problems associated with the actual application of the multispectral classification to the geoscience related fields. With the development of geographic information systems, more and more demands are required to speed up the processing of raw data interpretation and the renewal of the old information stored in the system, which is an essential effort to make a resource management system more efficient for the purpose of monitoring our changing environment.

One of the promising techniques to improve multispectral classification technology is the incorporation of information from different sources. When the information from different sources has some influence, either a substantial or subtle influence, on the development or the spectral behavior of the existing land covers, their combined impact on a classification scheme can result in a much better performance than that achieved by the classification scheme using only mono-source information.

In this thesis, the attention is focused on the possible improvement of classification performance by incorporating contextual information when the probabilistic relaxation procedure is applied to the urban environment. The problem with this procedure is the difficulty in finding and defining the information sources that have either direct or indirect impact on the spectral behavior of the urban land cover categories, because the urban areas are most likely to have developed under moderate ecological conditions. As
an alternative, the contextual information that does affect the behavior of a classification scheme is extracted from the numerical relationships existing among pixels and then is fed back into the classification scheme in a proper way to improve the overall performance of that scheme. This kind of information integration has been tried and tested on a urban scene using the procedure described in this thesis and its operation has resulted in an overall classification accuracy from 77.01% to 88.49% correct using a satellite image of 256 by 256 pixels.

The success of this procedure is twofold. First, the spectral separability checking has reduced greatly the spectral ambiguity that existed among the different land cover categories. By the operation of this step before classification, there is no substantial overlap of the spectral features in a 6-dimensional classification space, an essential requirement to the classification scheme that begins with the GML classification and followed by the probabilistic relaxation procedure to increase the accuracy. If such a step were not taken before the actual classification, we would not be able to know how much chance we have to make the classification errors, either by using the improperly selected training data or keeping the same spectral class as different informational classes, which eventually will jeopardize the fundamental assumptions required by the classification procedure. In other words, the proceeding with the classification scheme without knowing the spectral ambiguity is only in some sense a guessing game, especially when it is applied to the scene with higher heterogeneity in nature.

Second, when it is used on an urban environment, the probabilistic relaxation model is successful to summarize and apply the contextual information to improve the performance of the GML classifier. This information is contained in the neighborhood of a central pixel in the form of class membership probability and used in the form of class membership occurrence between different classes, or the compatibility coefficient, to adjust the probability of that central pixel's being one of the different classes. The
contextual information is used repeatedly in the iteration until the maximum accuracy is obtained. This important feature makes it possible for the relaxation model to reduce, even eliminate, continuously the identity ambiguity between different land cover categories. This identity ambiguity is caused primarily from the spectral ambiguity that we can not get rid of completely by the first step of the classification procedure.

An urban area represents a complicated system in terms of its pattern of development that in turn gives more complicated spectral and spatial patterns than those from agriculture and forest scenes, for example. After initial researchers had developed the relaxation model, it was first tested on the artificial imagery and then the real imagery with much homogeneity, but only few experiments have been done on the urban scene because of its heterogeneous nature in both spectral and spatial patterns. Though the relaxation model is successful when applied to the Omaha metropolitan area by the experiment conducted by the thesis research, it has also identified some problems that indicate the reason why it is not so impressive in improving the accuracy than those studies conducted by some researchers using different data sets (in some cases, the classification accuracy had reached as high as 95%).

It is found by the error pattern analysis that most errors have occurred at the boundaries between different land cover categories. These errors are partially eliminated and the class membership is adjusted by the relaxation model when it takes into account the contextual information to modify the probabilities, which has accounted for the major classification improvement made by the model. Some classification errors, however, occur inside the major category as individual parcels. From the theoretical considerations of the relaxation model itself, it is easier to adjust the error pixels at the boundaries because they are compatible to and have higher possibility to become the dominant categories at either side of the boundary. For those error pixels surrounded by a dominant category, however, they stay unchanged as error pixels if
these pixels occupy an area with an accountable geometric dimension and have a tendency of being incompatible to their background category. This is the limitation of the classification scheme itself, especially when it is operated on the scene with some heterogeneity.

6.2 Some Suggestions on the Further Research

A number of proposals can be made for further research. Firstly, a more quantitative approach to the spectral separability checking could be attempted. Instead of comparing their relative magnitudes of the statistical distances between different land cover categories, a threshold could possibly be set up to combine those categories with a statistical distance that is less than a specified magnitude. By setting up different levels of threshold, we may have a much more detailed idea on the risk of the pixel mis-assignment and the classification confidence in terms of possibility when using different classification schemes and data sets. When properly applying a statistical model to describe quantitatively these measurements, we can eliminate as much spectral ambiguity as possible to make multispectral classification more practical.

Secondly, when considering the spectral nature of the reflectance from different bands, it is found that a poor linear relationship exists between some bands that makes it difficult for any classification scheme to eliminate classification errors. As we are talking about the n-dimensional classification space, we have implied that the best result is produced from a theoretically ideal n-dimensional space in which each variable is independent of each other. In fact, in this data set, the independency between some specific bands is far from this requirement (for example, bands 2 and 3 of TM). This fact may hinder the performance of most classification schemes and sometimes may even
disable a classification scheme from functioning. To lesson the possible damage by the non-linearity of the data on the multispectral classification, a principle component analysis should be approached to improve the data linearity. After being transformed to their principle components, the data from different bands will become independent of each other and we can use only few of the major components to start a classification procedure to achieve a better result. The benefit of this approach is twofold: the accuracy and the speed. With the considerations of the statistical models discussed in Chapter 3, it is reasonably concluded that the data linearity can provide much better theoretical basis for those statistical models used in the thesis approach.

Thirdly, it is no question that a better result is expected when incorporating information from different sources into a classification system. The real, and difficult, question is how to define and format the useful information. Before any spatial, ancillary, and contextual information is integrated into a system, their real meanings to the environment under investigation have to be understood by the interpreter and their definitions have to be representative of the phenomena under study. Another problem we are facing is that we should be able to have the useful information formatted into a format that is compatible to the data to be used in the classification. For example, when we look at an image of an urban area, the spectral reflectance from the land covers is a universal phenomenon, while the different spatial patterns represented by the residential units with different compositions is only a local phenomenon. If this phenomenon is universal to all units with the same composition, it is valuable information to differentiate them from other regular units. Compared with the reflectance, however, it becomes the information with irregularity. For the purpose of the information integration, we need a paradigm to define and format this sort of information into meaningful and machine readable format to make the information integration operationable.
The last comments will address the problem of human and computer perceptions of the real world because some errors by the classification conducted by this thesis study come from the difference between two different types of perception. When a person reads a map interpreted manually, he or she will inevitably feel the unwillingness of the interpreter to include a land cover category into its background, especially when that particular category occupies certain geometric dimension, like the major streets in a city, or a small piece of grass in a residential area. This subtle interpretation is required in large scale mapping, but is not visually appealing to a small scale mapping using satellite data. The computer perception of the world, on the other hand, is the generalization of different land cover categories by the spectral features rather than by their actual functions. The network of the geographical features, like the streets in a metropolitan area, can be overlaid on the top of the machine interpreted map to make it more like the human perception of that area. The more detailed is a classification attempted by machine, the more is it subjected to errors, and the more intelligence is needed by the machine. How much those two types of perception could be matched and how much the machine intelligence could be exploited will be a research topic for many years to come.
APPENDICES

PROGRAM LISTINGS

1. PROGRAM REFORMAT 109
2. PROGRAM DIVERGENCE 111
3. PROGRAM J-M DISTANCE 114
4. PROGRAM GROUND TRUTH STATISTICS 117
5. PROGRAM WATER BOUNDARY 118
6. PROGRAM GML 119
7. PROGRAM CONVERSION OF PROBABILITY TO MAP 123
8. PROGRAM WATER BODY PRINTING 125
9. PROGRAM ACCURACY COMPARISON 126
10. PROGRAM COMPATIBILITY COEFFICIENT 128
11. PROGRAM UPDATING PROBABILITIES 132
PROGRAM REFORMAT

c This is a program to reformat a signature file from ERDAS program to the
c format that is acceptable by other programs. The input is from a
c six-band-ten-category training data, named sig.dat

program reformat
  dimension comat(6),cmean(6,10),std(6,10)
  character*80 title2,title3
  character*10 cofile,mnstd

c Enter file name from screen
  write(6,100)
100  format('$Enter name of covariance matrix file: ')
  read(6,'(a10)')cofile
  write(6,110)
110  format('$Enter name of mean value and standard deviation file: ')
  read(6,'(a10)')mnstd
  open(unit=1,file='sig.dat',status='old',err=500)
  open(unit=2,file=cofile,status='new')
  open(unit=3,file=mnstd,status='new')
  read(1,'(a80)')title2
  read(1,'(a80)')title3
  goto 5

c Print error message
500  write(6,*)
    write(*>('File sig.dat not exist, check the file name or have it
    x created by ERDAS program!'))
    goto 400

c Read heading from input data and write headings to the output files
5   read(1,*nb,nc
   write(2,'(x,a80)')title2
   write(2,*
   write(2,*nb,nc
   write(3,'(x,a80)')title3
   write(3,*
   write(3,*nb,nc

c Start main loop
    do 10 i=1,nc
      read(1,'(////)')

c Read mean values and standard deviations for one category
    do 20 j=1,nb
      read(1,'(10x,f8.3,9x,f8.3)')cmean(j,i),std(j,i)
20   continue
      read(1,'(/////)')
c Read covariance matrix of one category and write it back to the output file
  do 30 j=1,nb
    read(1,'(7x,<nb>f8.3)')(comat(k),k=1,nb)
    write(2,'(<nb>f11.5)')(comat(k),k=1,nb)
  30 continue

c Append the mean value to the last line of covariance from one category
  write(2,'(<nb>f11.5)')(cmean(j,i),j=1,nb)
  write(2,'(/)')
  10 continue

c End of main loop
  close(1)
  close(2)

c Write mean values and standard deviations to another file for documentation
  do 40 i=1,nb
    write(3,'(<nc>f8.3)')(cmean(i,j),j=1,nc)
  40 continue
  write(3,'(/)')
  do 50 i=1,nb
    write(3,'(<nc>f8.3)')(std(i,j),j=1,nc)
  50 continue
  close(3)

400 end
PROGRAM DIVERGENCE

c This a program to calculate divergence between different classes using
c a 6-band-ten-category training data and create the comparison table

program divergence
character*16 class, compar
dimension comat(6,6,10), cinvm(6,6,10), avg(10,6), tesres(6,6)
dimension avg1(6), avg2(6), resmid(6,6), produt(6,6), comp(10)

c Enter input/output file names from screen
109 write(*,110)
110 format('$Enter the name of the original data file: ')
read('*','(a16)') class
open(unit=1 ,file=class,status='old',err=500)
write(*,120)
120 format('$Enter name of comparison matrix file: ')
read('*','(a16)') compar
go to 5

c Print error message on screen
500 write(*,*)
write(*,'*')'File not found, try again!'
write(*,*)
go to 109

c Training data input
5 read(1,('/'))
read(1,'*)&nb,nc
do 10 k=1,nc
do 12 i=1,nb
read(1,'(1x,<nb>f11.5)')(comat(i,j,k),j=1,nb)
do 12 j=1,nb
cinvm(i,j,k)=comat(i,j,k)
continue
read(1,'(1x,<nb>f11.5)')(avg(k,j),j=1,nb)
read(1,('/'))
10 continue
close(1)

c Calculate the inverse of matrix
do 13 kcat=1,nc
call gjdef(nb,nc,kcat,cinvm)
13 continue
open(unit=2 ,file=compar,status='new')
write(2,'(3x,<nc>i8)')(i,i=2,nc)
write(2,('(i)')')

c Calculate divergence
do 70 kcat=1,nc-1
do 60 kcat1=kcat+1,nc
call matadd(nb,nc,kcat,kcat,1,cinvm,tesres,1.0)
do 20 i=1,nb
   avg1(i)=avg(kcat,i)-avg(kcat,1)
20 continue
do 30 i=1,nb
   add=0.0
   do 40 j=1,nb
      add=tesres(i,j)*avg1(j)
40 continue
   avg2(i)=add
30 continue
do 50 i=1,nb
   do 50 j=1,nb
      tesres(i,j)=avg2(i)*avg1(j)
50 continue
call trace(nb,tesres,dig)
acctra=dig
call matadd(nb,nc,kcat,1,kcat,cinvm,tesres,-1.0)
call matadd(nb,nc,kcat,kcat,1,comat,resmid,-1.0)
call matmul(resmid,tesres,produt,nb,nb,nb)
call trace(nb,produt,dig)
comp(kcat)=acctra+dig)/2.0
60 continue

c Output divergence matrix
write(2,65)kcat,(int(comp(i)*10.0+0.5),i=kcat+1,nc)
65 format(1x,i2,<\(kcat-1\)\)*8+1>x,<nc-kcat>i8)
70 continue
close(2)

c This the subroutine to calculate the inverse of matrix

subroutine gjdef(mb,mc,kct,a)
dimension a(mb,mb,mc)
do 20 k=1,mb
   if(abs(a(k,k,kct)).lt.0.00001)goto 50
   q=a(k,k,kct)
do 15 i=1,mb
      a(k,i,kct)=a(k,i,kct)/q
15 continue
do 30 i=1,mb
   if(i.eq.k)goto 30
   p=-a(i,k,kct)
do 40 j=1,mb
      a(i,j,kct)=a(k,j,kct)*p+a(i,j,kct)
40 continue
   a(i,k,kct)=p/q
30 continue
a(k,k,kct)=1/q
20 continue
return
50 write(6,*)k,a(k,k,kct)
write(6,60)
60 format('fail in gjdef - matrix singular')
end

c This is the subroutine to multiply two matrixes

subroutine matmul(a,b,c,m1,m2,m3)
dimension a(m1,m2),b(m2,m3),c(m1,m3)
do 10 i=1,m1
do 10 j=1,m3
c(i,j)=0.0
10 continue
do 30 k=1,m1
do 30 j=1,m3
do 30 i=1,m2
    c(k,j)=a(k,i)*b(i,j)+c(k,j)
30 continue
end

c This is the subroutine to calculate the trace of a matrix

subroutine trace(mb,a,p)
dimension a(mb,mb)
p=0.0
do 10 i=1,mb
    p=a(i,i)+p
10 continue
end

c This is the subroutine to calculate sum/difference of two matrixes

subroutine matadd(mb,mc,k,k1,a,c,b)
dimension a(mb,mb,mc),c(mb,mb)
do 20 i=1,mb
do 20 j=1,mb
    c(i,j)=a(i,j,k)+b*(a(i,j,k1))
20 continue
end
PROGRAM J-M DISTANCE

c This is a program to calculate J-M distance between different classes
c using a six-band-ten-category training data and create comparison table

    program JM distance
    character*16 class,compar
    dimension comat(6,6,10),avg(10,6),comp(10),dtm(10)
    dimension tesres(6,6),avag1(6),avag2(6),resmid(6,6)

c Enter input/output filenames from screen
109     write('*,110)
110     format('$Enter the name of original data file: ')
         read('*,'(a16))')class
         open(unit=1,file=class,status='old',err=500)
         write(*,120)
120     format('$Enter the name of comparison matrix file: ')
         read('*,'(a16))')compar
         goto 5

c Print error message on screen
500     write('*,*)
         write('*,*)'File not found, try again!'
         write('*,*)
         goto 109

c Training data input
5     read(1,'(/)')
     read(1,'(*)')nb,nc
     do 12 k=1,nc
     do 10 i=1,nb
     read(1,'(1x,<nb>f11.5)')(comat(i,j,k),j=1,nb)
10     continue
     read(1,'(1x,<nb>f11.5)')(avg(k,j),j=1,nb)
     read(1,'(/)')
12     continue
     close(1)

c Calculate the determinants of all covariance matrices
     open(unit=2,file=compar,status='new')
     write(2,'(3x,<nc)i8)')(i,i=2,nc)
     write(2,'(/)')
     do 15 kcat=1,nc
     do 17 i=1,nb
     do 17 j=1,nb
     tesres(i,j)=comat(i,j,kcat)
17     continue
     call determ(nb,nc,kcat,tesres,dtm)
15     continue

c Start main loop
do 18 kcat=1,nc-1
do 16 kcat1=kcat+1,nc
do 20 i=1,nb

c Start calculating the coefficient
    avag1(i)=avg(kcat,i)-avg(kcat1,i)
20 continue
    call matadd(nb,nc,kcat,kcat1,comat,resmid)
    call matadd(nb,nc,kcat,kcat1,comat,tesres)
call determ(nb,nc>0,tesres,dtm)
c Calculate none-quadratic term of the coefficient
    bdd=sqrt(dtm(kcat)*dtm(kcat1))
    bdd=(log(dtm(0)/bdd))/2.0

c Calculate quadratic term of the coefficient
    call gjdef(nb,resmid)
do 30 i=1,nb
    add=0.0
    do 40 j=1,nb
        add=avag1(j)*resmid(j,i)+add
40 continue
    avag2(i)=add
30 continue
do 50 i=1,nb
    add=avag1(i)*avag2(i)+add
50 continue
    add=add/8.0+bdd

c Calculate the J-M distance using the coefficient
    comp(kcat1)=sqrt(2.0*(1.0-exp(-add/100.0)))
16 continue

c Write to the output comparison table
write(2,90)kcat,(int(comp(i)*1000+0.5),i=kcat+1 ,nc
90 format(1x,i2,<(kcat-1)*8+1>x,<nc-kcat>i8)
18 continue

c End of main loop
    close(2)

end

c This is the subroutine to calculate the inverse of a matrix

subroutine gjdef(m,a)
dimension a(m,m)
do 20 k=1,m
    if(abs(a(k,k)).lt.0.0000001) goto 50
q = a(k,k)
do 15 i=1,m
   a(k,i) = a(k,i)/q
15 continue
do 30 i=1,m
   if(i.eq.k)goto 30
   p = -a(i,k)
do 40 j=1,m
   a(i,j) = a(k,j)*p + a(i,j)
40 continue
   a(i,k) = p/q
30 continue
   a(k,k) = 1/q
20 continue
return
50 write(6,60)
60 format('fail in gjdef - matrix singular')
end

This is the subroutine to calculate the sum of two matrixes

subroutine matadd(mb,mc,k1,a,c)
dimension a(mb,mb,mc),c(mb,mb)
do 10 i=1,mb
   do 10 j=1,mb
      c(i,j) = (a(i,j,k) + a(i,j,k1))/2.0
   c write(6,*)c(i,j),a(i,j,k),a(i,j,k1),i,j
10 continue
end

This is the subroutine to calculate the determinant of a matrix

subroutine determ(mb,mc,kct,a,dp)
dimension a(mb,mb),dp(mc)
do 10 k=1,mb-1
   do 10 i=k+1,mb
      dp(kct) = -a(i,k)
   do 10 j=k,mb
      a(i,j) = a(k,j)*dp(kct)/a(k,k) + a(i,j)
10 continue
   dp(kct) = 1.0
    do 20 i=1,mb
       dp(kct) = dp(kct)*a(i,i)
20 continue
end
PROGRAM GROUND TRUTH STATISTICS

c This is a program to create ground truth statistics

program ground truth statistics
dimension ima(256),icat(6)
integer*4 icat
byte ima
character*16 grfile,icfile

c Enter input/output file from screen
99 write(*,100)
100 format('Enter name of ground truth classification file: ')
    read(*,'(a16)')grfile
    open(unit=1,file=grfile,form='unformatted',status='old',err=500)
    write(*,110)
110 format('Enter name of category statistics file: ')
    read(*,'(a16)')icfile
    open(unit=2,file=icfile,status='new')
goto 5

c Print error message on screen
500 write(*,*)
    write(*,'File not found, try again')
    write(*,*)
goto 99
5 nb=6
nc=6
isize=256
write(*,'(////)')

c Read one line of the image and calculate statistics from the ground truth data
do 10 i=1,isize
    read(1)(ima(j),j=1,isize)
    write(*,200)i
200 format('+',20x,'The current line number is: ',i3)
    do 20 j=1,isize
        icat(ima(j)) = icat(ima(j)) + 1
    20 continue
10 continue
close(1)

c Write statistics to a file
    write(2,'Category statistics')
    write(2,*)
    write(2,'(1x,a11,<nc>i10)')'Category #:',(i,i=1,nc)
    write(2,'(1x,a11,3x,<nc>i10)')'Pixel #:',(icat(i),i=1,nc)
    write(2,300)'Percentage: ',((real(icat(i))/real((isize)**2.0)),i=1,nc)
300 format(1x,a11,3x,<nc>f10.6)
close(2)
end
PROGRAM WATER BOUNDARY

c This is a program to output the water body extent in terms of numbers of beginning and ending pixels

    program water body limit
    dimension nwater(256),iden(2)
    character*16 water
    byte nwater

c Enter input/output file names
99  write(*,100)
100  format('Enter name of ground truth file: ') water
    read(*,'(a16)') water
    open(unit=1,file=water,form='unformatted',status='old',err=500)
goto 5

c Print error message on screen
500  write(*,*)
    write(*,*)'File not found, try again!' write(*,*)
goto 99

5  open(unit=2,file='water.out',status='new')
isize=256
    do 10 i=1,isize
    read(1)(nwater(j),j=1,isize)
icount=0
    do 20 j=1,isize
      if(nwater(j).eq.2)then
        iden(1)=i
        iden(2)=j-icount
        icount=icount+1
      endif
    20 continue
    if(icount.ne.0)then
      write(2,'(1x,3|5)')iden(1),iden(2),icount
    endif
    continue
10  close(1)
close(2)
end
PROGRAM GML

This is a program to create a matrix of category probability.
This matrix consists of 256 by 256 by 6 elements, where 256 is
the size of imagery and 6 is the number of category. User can
change the size to fit his/her applications.

```
program gml
  dimension ima(6,256),pro(6,256),comat(6,6,6),comean(6,6),de(6)
dimension tesres(6,6),image(6,256)
byte image
character*16 imfile,prfile,cofile

! Enter input/output file names from screen
129 write(*,130)
130 format('Enter name of the image file: ')
   read(*,'(a16)')imfile
   open(unit=1,file=imfile,form='unformatted',status='old',err=500)
149 write(*,150)
150 format('Enter name of the covariance and mean value matrix file: ')
   read(*,'(a16)')cofile
   open(unit=3,file=cofile,status='old',err=520)
   write(*,140)
140 format('Enter name of the probability file: ')
   read(*,'(a16)')prfile
   open(unit=2,file=prfile,form='unformatted',status='new')
goto 5

! Print error messages on the screen
500 write(*,*)
   write(*,'File not found, try again!')
   write(*,*)
goto 129
520 write(*,*)
   write(*,'File not found, try again!')
   write(*,*)
goto 149

! Read mean value and covariance matrices into matrix comat
5   read(3,'(/)')
   read(3,'*')nb,nc
   isize=256
   do 10 k=1,nc
      do 15 i=1,nb
         read(3,'(1x,<nb>f11.6)')(comat(i,j,k),j=1,nb)
      15 continue
   read(3,'(/)')(comean(k,j),j=1,nb)
   read(3,'(/)')
10 continue
   close(3)
```
c Apply logarithmic transformation to the probability density function
c and calculate the terms that are related only to covariance matrices

    do 20 kcat=1,nc
    do 16 i=1,nb
    do 16 j=1,nb
        tesres(i,j)=comat(i,j,kcat)
    continue
    call determ(nb,nc,kcat,tesres,de)
    call gjdef(nb,nc,kcat,comat)
    20 continue

do 30 i=1,nc
    de(i)=log(2*3.1415926)*nb/2.0+log(de(i))/2.0
30 continue

c Write heading to the probability file
    write(2)'This is the probability file'
    write(2)
    write(2)nb,nc,isize
    write(*,'(///)')

c Start the main loop
    do 40 line=1,isize
        write(* ,200) line
        200 format('+',20x,'The current line number is: ',i3)

c Read one line of the image, each line consisting of 6-band data
    do 45 i=1,nb
        read(1)(image(i,j),j = 1,isize)
45 continue

c Transfer the integer image into the byte image
    do 33 i=1,nb
        do 33 j=1,isize
            if(image(i,j).lt.0) then
                ima(i,j)=image(i,j)+256
            else
                ima(i,j)=image(i,j)
            endif
43 continue

c Compute the quadratic term of the logarithmic probability
    call gmlpr(nb,nc,isize,ima,profcomat,comean)

c Start scaling and normalization of the probability
    do 35 j=1,isize
        sum=0.0
35 continue

c Transfer logarithmic probability into real probability
    pro(i,j)=exp((-pro(i,j)-de(i))/100.0)
    sum=sum+pro(i,j)
continue
do 35 i=1,nc

c Normalization
pro(i,j)=pro(i,j)/sum
continue

c Data output
c Write one line of probability to the probability file, each line
c consisting of 6-category probabilities
do 40 lpr=1,nc
write(2)(pro(ipr,js),js=1,isize)
continue

c End of main loop
close(1)
close(2)
end

This is the subroutine to compute the qprobability density function
This from one line image

subroutine gmlpr(mb,mc,isz,im,pr,cim,cm)
dimension im(mb,isz),pr(mc,isz),cim(mb,mb,mc),cm(mc,mb),av(6)
do 10 j=1,isz
do 10 k=1,mc
do 20 i=1,mb
av(i)=real(im(i,j))-cm(k,i)
continue
as=0.0
do 30 jj=1,mb
af=0.0
do 40 ii=1,mb
af=af+av(ii)*cim(ii,jj,k)
continue
as=as+af*av(jj)
continue
pr(k,j)=as/2.0
continue
end

This is the subroutine to create the inverse of matrix

subroutine gjdef(mb,mc,kct,a)
dimension a(mb,mb,mc)
do 20 k=1,mb
if(abs(a(k,k,kct)).lt.0.00000001)goto 50
q=a(k,k,kct)
do 15 i=1,mb
  a(k,i,kct)=a(k,i,kct)/q
15  continue
do 30 i=1,mb
  if(i.eq.k)goto 30
  p=-a(i,k,kct)
do 40 j=1,mb
  a(i,j,kct)=a(k,j,kct)*p+a(i,j,kct)
40  continue
a(i,k,kct)=p/q
30  continue
a(k,k,kct)=1.0/q
20  continue
return
50  write(*,*)'Fail in gjdef - matrix singular'
end

This is the subroutine to calculate the determinant of a matrix

subroutine determ(mb,mc,kct,a,p)
dimension a(mb,mb),p(mc)
do 10 k=1,mb-1
do 10 i=k+1,mb
  p(kct)=-a(i,k)
do 10 j=k,mb
  a(i,j)=a(k,j)*p(kct)/a(k,k)+a(i,j)
10  continue
  p(kct)=1.0
do 20 i=1,mb
  p(kct)=p(kct)*a(i,i)
20  continue
end
PROGRAM CONVERSION OF PROBABILITY TO MAP

c This is a program to convert probability to actual land use category

c using the maximum likelihood decision rule

    program probability to map
    dimension p(6,256),icat(256),conpro(6)
    byte icat
    character*16 prfile,cafile

c Enter input/oupt file names from screen
    write(*,100)
    100 format('$Enter name of the probability file: ')
            read(*,'(a16)')prfile
            open(unit=1,file=prfile,form='unformatted',status='old',err=500)
            write(*,120)
    120 format('$Enter name of landuse category file: ')
            read(*,'(a16)')cafile
            goto 5

c Print error message on screen
    500 write(*,*)
            write(*,'file not found, try again!')
            write(*,*)
            goto 99

c Skip the heading of the probability file
    5 open(unit=2,file=cafile,form='unformatted',status='new')
        read(1)
        read(1)
        read(1)nb,nc,isize

c Write the heading to the land cover map file
    write(2)'Landuse map of Omaha'
    write(2)
    write(2)nb,nc,isize
    write(*,'(////////)')

c Start converting probability file to land cover map file
    do 10 ii=1,isize
        write(*,150)ii
    150 format('+',20x,'The current line number is: ',i3)
            do 20 i=1,nc

c Read one line of probabilities, each line consisting of 6-class probabilities
    read(1)(p(i,j),j=1,isize)
    20 continue

c Find the maximum probability
    do 30 j=1,isize
a = p(1, j)
max = 1
do 50 i = 2, nc
b = p(i, j)
if(b > a) then
   a = b
   max = i
endif
50 continue
icat(j) = max
30 continue

c Write one line land cover categories to the land cover map
write(2)(icat(i), i = 1, isize)
10 continue

c End of loop
   close(1)
   close(2)

end
PROGRAM WATER BODY PRINTING

c This is a program to print out the water extent after each iteration

    program water extent
    dimension nwater(256)
    byte nwater
    character*16 iwater, owater

    c Enter input file name
    69   write('*,70)
    70   format('Enter name of land cover map: ')
        read('*(a16)',iwater)
        open(unit=1, file=iwater, form='unformatted', status='old', err=400)
        goto 5

    c Print error message on screen
    400 write('**')
        write('**','File not found, try again!')
        write('**')
        goto 69

    c Read land cover map
    5   open(unit=2, file='water.out', status='old')
        read(1)
        read(1)
        read(1) nb, nc, isize
        write('**',50)
        50   format('Enter name of file for water coverage: ')
            read('*(a16)',owater)
            open(unit=3, file=owater, status='new')
            i = 1

    c Read water body limit in terms of the numbers of beginning and ending pixels
    100  read(2,'(1x,3i5)',end=500) irow, icol, npix
    200  read(1)(nwater(j), j=1, isize)
        if(i.eq.irow)then
            write(3,'(<70-icol>x,<npix>i1)')(nwater(j), j=icol, icol+npix-1)
            i = i+1
            goto 100
        else
            i = i+1
            goto 200
        endif
    500  close(1)
    close(2)
    close(3)

    end
PROGRAM ACCURACY COMPARISON

c This is a program to check the accuracy of classification and
create the error matrix for comparison

program check accuracy
dimension igtru(256),icatup(256),class(0:7),merror(7,7)
character*16 grfile,upfile,sufile
character*9 class
byte igtru,icatup

data class"COMIN','WATER','GRASS','TREES',
1 'BAREGR','RESID','TOTALS'/
c Enter input/output file names from screen
99 write(*,100)
100 format('$Enter name of ground truth file: ') read('*','(a16)')grfile
open(unit=1,file=grfile,form='unformatted',status='old',err=500)
109 write(*,110)
110 format('$Enter name of the updated classification map file: ') read('*','(a16)')upfile
open(unit=2,file=upfile,form='unformatted',status='old',err=510)
write(*,130)
130 format('$Enter name of the error matrix: ') read('*','(a16)')sufile
goto 5

c Print error message on screen
500 write(*,*)
write(*,*)'File not found, try againl'
write(*,*)
goto 99
510 write(*,*)
write(*,*)'File not found, try againl'
write(*,*)
goto 109

c Skip the heading of updated classification map
5 read(2)
read(2)nb,nc,isize
write(*,'(////)')
c Start main loop
   do 10 i=1,isize
      write(*,150)i
150 format('+',20x,'The current line number is: ',i3)
c Read one line of data from both ground truth map and the updated
c classification map
read(1)(igtru(j),j=1,isize)
read(2)(icatup(j),j=1,isize)
do 10 j=1,isize
  merror(igtru(j),icatup(j))=merror(igtru(j),icatup(j))+1
10 continue
c End of loop
  close(1)

c Delete the previous updated classification map, only the current
classification map is kept for the next iteration
  close(unit=2,status='delete')

c Compile error matrix by comparing current classification map with
c the ground truth map
  open(unit=10,file=sufile,status='new')
  write(10,*)'Error matrix'
  write(10,('(a9)'))
  do 20 i=1,nc
    do 30 j=1,nc
      merror(i,nc+1)=merror(i,nc+1)+merror(i,j)
      merror(nc+1,j)=merror(nc+1,j)+merror(i,j)
    30 continue
    merror(nc+1,nc+1)=merror(nc+1,nc+1)+merror(i,i)
  20 continue
  write(10,(1x,a17,7x,a19,12x,a7))'ERROR OF COMMISSION','ERROR OF OMISSION','CORRECT'
class(nc+1)='OVERALL'
do 50 i=1,nc
  if(merror(i,nc+1).ne.0)then
    mom=merror(i,nc+1)-merror(i,i)
    mco=merror(nc+1,i)-merror(i,i)
    mtl=merror(i,nc+1)
    write(10,(1x,a8,2i6,f7.2,7x,2i6,f7.2,6x,2i6,f7.2))class(i),
      mom,mtl,real(mom*100)/real(mtl),mco,mtl,real(mco*100)/real(mtl),
  1 iersum=iersum+mom
  endif
50 continue
write(10,(1x,a8,2i6,f7.2,7x,2i6,f7.2,6x,2i6,f7.2))class(nc+1),
  iersum,isize*isize,real(iersum*100)/real(isize*isize),iersum,
  isize*isize,real(iersum*100)/real(isize*isize),merror(nc+1,nc+1),
  isize*isize,real(merror(nc+1,nc+1)*100)/real(isize*isize)
close(10)
end
PROGRAM COMPATIBILITY COEFFICIENT

This is a program to create a compatibility coefficient matrix from probability file

program compatibility coefficient
dimension pij(3,256,6),rij(6,6,9),pca(6)
character*16 prfile,rfile
character*8 coef

Enter input/output file names from screen
119 write(*,120)
120 format('$Enter name of probability file: ')
read(*,'(a16)')prfile
open(unit=1,file=prfile,form='unformatted',status='old',err=500)
write(*,130)
130 format('$Enter name of compatability coeffient file ')
read(*,'(a16)')rfile
goto 5

Print error message on screen
500 write(*,*)
write(*,'File not found, try again!')
write(*,*)
goto 119

Read numbers of band, category and size of the imagery
5 read(1)
read(1)
read(1)nb,nc,isize

Read the first two lines of the probability, each line consisting of 6-class probability
10 do i=1,2
10 do k=1,nc
read(1)(pij(i,j,k),j=1,isize)
10 do j=1,isize
pca(k)=pca(k)+pij(i,j,k)
10 continue
write(*,'(////)')

Account the contribution of those pixels at the upper left and upper right corners and the upper horizontal edge to the construction of the compatibility coefficient matrix
10 call corner(pij,rij,isize,1,1,5,6,8,9,nc)
10 call corner(pij,rij,isize,1,isize,isize-1,4,5,7,8,nc)
10 call edgeh(pij,rij,isize,1,4,5,6,7,8,9,nc)

Start main loop
30 do ii=3,isize
write(*,150)ii

128
do 40 k=1,nc
  c Read one line of probability and proceed to calculate compatibility coefficients
  read(1)(pij(3,j,k), j=1,isize)
  do 40 j=1,isize
    pca(k)=pca(k)+pij(3,j,k)
  40 continue
  c Account the contribution of those pixels at the left edge to the construction of the compatibility coefficient matrix
  call edgev(pij,rij,isize,1,1,2,3,5,6,8,9,nc)
  c Account the contribution of those pixels at not located at the edges and corners to the construction of the compatibility coefficient matrix
  call comcof(pij,rij,isize,nc)
  c Account the contribution of those pixels at the right edge to the construction of the compatibility coefficient matrix
  call edgev(pij,rij,isize,isize,isize-1,1,2,4,5,7,8,nc)
  c Move the last two lines of probability to the first two lines of the 3 by 256 by 6 matrix
  do 50 k=1,nc
    do 50 j=1,isize
      pij(1,j,k)=pij(2,j,k)
      pij(2,j,k)=pij(3,j,k)
  50 continue
  30 continue
  c End of main loop
  close(1)
  c Account the contribution of those pixels at the lower left and lower right corners and the lower horizontal edge to the construction of the compatibility coefficient matrix
  call corner(pij,rij,isize,2,1,1,1,2,3,5,6,nc)
  call corner(pij,rij,isize,2,isize,isize-1,1,2,4,5,nc)
  call edgeh(pij,rij,isize,2,1,2,3,4,5,6,nc)
  c Make the compatibility coefficients in the range of [-1,1]
  rmax=0.1
  do 60 k=1,9
    do 60 i=1,nc
      do 60 j=1,nc
        rrij(i,j,k)=log(isize**2*rij(i,j,k)/((pca(i)*pca(j))))
        if(abs(rij(i,j,k)).gt.abs(rmax))then
          rmax=rij(i,j,k)
        endif
      60 continue
    60 continue
  60 continue
  do 65 k=1,9
    do 65 i=1,nc

129
do 65 j=1,nc
rij(i,j,k)=rij(i,j,k)/abs(rmax)
65 continue
c Write compatibility coefficients to a file
open(unit=2,file=rfile,status='new')
write(2,21)
21 format(1x,'Compatibility coefficient file')
write(2,*)
coef='Rij(k,k)'
do 70 k=1,9
coef(5:5)=char(48+k)
coef(7:7)=char(48+k)
write(2,'(1x,a8)')coef
do 70 i=1,nc
write(2,'(1x,<nc>f10.6)')(rij(i,j,k),j = 1,nc)
70 continue
write(2,*)
close(2)
end
c This is the subroutine to calculate compatibility coefficients at corner
subroutine corner(p,r,isz,l1,l2,l3,k1,k2,k3,k4,mc)
dimension p(3,isz,mc),r(mc,mc,9)
do 10 ll=1,mc
a=p(ll,1,li)
do 10 lr=1,mc
r(ll,lr,k1)=r(ll,lr,k1)+p(1,l3,lr)*a
r(ll,lr,k2)=r(ll,lr,k2)+p(1,l3+1,lr)*a
r(ll,lr,k3)=r(ll,lr,k3)+p(2,l3,lr)*a
r(ll,lr,k4)=r(ll,lr,k4)+p(2,l3+1,lr)*a
10 continue
end
c This is the subroutine to calculate compatibility coefficients at edge-
c horizontal
subroutine edgeh(p,r,isz,l1,k1,k2,k3,k4,k5,k6,mc)
dimension p(3,isz,mc),r(mc,mc,9)
do 10 j=2,isz 1
do 10 ll=1,mc
a=p(ll,j,li)
do 10 lr=1,mc
r(ll,lr,k1)=r(ll,lr,k1)+p(1,j-1,lr)*a
r(ll,lr,k2)=r(ll,lr,k2)+p(1,j,lr)*a
10 continue

\[ r(ll, lr, k3) = r(ll, lr, k3) + p(1, j+1, lr) \cdot a \]
\[ r(ll, lr, k4) = r(ll, lr, k4) + p(2, j-1, lr) \cdot a \]
\[ r(ll, lr, k5) = r(ll, lr, k5) + p(2, j, lr) \cdot a \]
\[ r(ll, lr, k6) = r(ll, lr, k6) + p(2, j+1, lr) \cdot a \]

continue

end

c This is the subroutine to calculate compatibility coefficients at edge-vertical

subroutine edgev(p, r, isz, M, k1, k2, k3, k4, k5, k6, mc)
dimension p(3, isz, mc), r(mc, mc, 9)

do 10 ll=1, mc
   a=p(2, ll, ll)
do 10 lr=1, mc
      r(ll, lr, k1) = r(ll, lr, k1) + p(1, l2, lr) \cdot a
      r(ll, lr, k2) = r(ll, lr, k2) + p(1, l2+1, lr) \cdot a
      r(ll, lr, k3) = r(ll, lr, k3) + p(2, l2, lr) \cdot a
      r(ll, lr, k4) = r(ll, lr, k4) + p(2, l2+1, lr) \cdot a
      r(ll, lr, k5) = r(ll, lr, k5) + p(3, l2, lr) \cdot a
      r(ll, lr, k6) = r(ll, lr, k6) + p(3, l2+1, lr) \cdot a
   10 continue
end

c This is the subroutine to calculate every nine compatibility coefficients using class probabilities associated with the neighboring pixels of a central pixel

subroutine comcof(p, r, isz, mc)
dimension p(3, isz, mc), r(mc, mc, 9)

do 10 j=2, isz-1
do 10 k=1, mc
   a=p(2, j, k)
do 10 k1=1, mc
      r(k, k1, 1) = r(k, k1, 1) + p(1, j-1, k1) \cdot a
      r(k, k1, 2) = r(k, k1, 2) + p(1, j, k1) \cdot a
      r(k, k1, 3) = r(k, k1, 3) + p(1, j+1, k1) \cdot a
      r(k, k1, 4) = r(k, k1, 4) + p(2, j-1, k1) \cdot a
      r(k, k1, 5) = r(k, k1, 5) + p(2, j, k1) \cdot a
      r(k, k1, 6) = r(k, k1, 6) + p(2, j+1, k1) \cdot a
      r(k, k1, 7) = r(k, k1, 7) + p(3, j-1, k1) \cdot a
      r(k, k1, 8) = r(k, k1, 8) + p(3, j, k1) \cdot a
      r(k, k1, 9) = r(k, k1, 9) + p(3, j+1, k1) \cdot a
   10 continue
end
PROGRAM UPDATING PROBABILITIES

c This is a program to update probabilities using contextual information contained in the neighborhood of a central pixel

    program update
    dimension pij(3,256,6),rij(6,6,9),p(6,256),q(6)
    character*16 prfile,upfile,rfile

    c Enter input/output file names from screen
    99   write(*,100)
    100  format('Enter name of probability file from previous iteration: ')
    read(*,(a16))prfile
    open(unit=1,file=prfile,form='unformatted',status='old',err=500)
    109  write(*,110)
    110  format('Enter name of compatibility coefficient file: ')
    read(*,(a16))rfile
    open(unit=2,file=rfile,status='old',err=510)
    119  write(*,120)
    120  format('Enter name of new probability file after current iteration: ')
    read(*,(a16))upfile
    goto 5

    c Print error message on screen
    500  write(*,*)
         write(*,*)'File not found, try again!'
         write(*,*)
         goto 99
    510  write(*,*)
         write(*,*)'File not found, try again!'
         write(*,*)
         goto 109

    c Skip heading and read the numbers of band, category and size of imagery
    5    read(1)
         read(1)
         read(1)nb,nc,isize

    c Read compatibility coefficients
    10   do 20 k=1,9
         do 10 i=1,nc
         read(2,'(1x,<nc>f10.6)')(rij(i,j,k),j=1,nc)
    20   continue
    10   continue
    write(2,*)
    close(2)

    c Write heading to the updated probability file
open(unit=3,file=upfile,form='unformatted',status='new')
write(3)'Updated probability file'
write(3)
write(3)nb,nc,isize

c Read first two lines of probability, each line consisting of 6-class probability
do 11 i=1,2
do 11 k=1,nc
read(1)(pij(i,j,k),j=1,isize)
continue

11
c Update those probabilities located at the upper left and upper right corners and the upper horizontal edge
qmin=100.0
call upcorner(pij,rij,q,p,nc,isize,1,1,0,5,6,8,9,qmin)
call upcorner(pij,rij,q,p,nc,isize,1,isize-1,1,4,5,7,8,qmin)
call upedgeh(pij,rij,q,p,nc,isize,1,4,5,6,7,8,9,qmin)

c Write the first line of updated probabilities to the updated probability file
do 13 i=1,nc
write(3)(p(i,j),j=1,isize)
continue
write(*,'(////)')

13
c Start main loop
do 27 line=3,isize
write(*,150)line
150 format('+',20x,'The current line number is: ',i3)

c Read the third line of the probability
do 30 k=1,nc
read(1)(pij(3,j,k),j=1,isize)
continue

c Update the probabilities located at the left and right edges
call upedgev(pij,rij,q,p,nc,isize,1,0,2,3,5,6,8,9,qmin)
call upedgev(pij,rij,q,p,nc,isize,isize-1,1,2,4,5,7,8,qmin)

c Update probabilities not located at the edges by moving a 3 by 3 window from left to right
do 40 j=2,isize-1
do 50 i=1,nc
q(i)=0.0
50 continue
do 60 ii=1,nc
do 60 jj=1,nc
do 60 id=1,3
md=3*id-1
do 60 jd=j-1,j+1
q(ii)=q(ii)+rij(ii,jj,md+jd-j)*pij(id,jd,jj)
60 continue
a=0.0
do 70 i=1,nc
 a=a+pij(2,j,i)*(7.0+q(i))
if(q(i).lt.qmin)then
  qmin=q(i)
endif
70 continue

do 40 i=1,nc
 p(i,j)=pij(2,j,i)*(7.0+q(i))/a
40 continue

write one line of the updated probabilities to the output file

write(3)(p(i,j),j=1,isize)

move last two lines of probabilities to the first two lines

do 90 k=1,nc
do 90 j=1,isize
pij(1,j,k)=pij(2,j,k)
pij(2,j,k)=pij(3,j,k)
90 continue

end of main loop

delete the previous probability file, only the current probability file
c is kept for the next iteration

close(unit=1,status='delete')

update the probabilities located at the lower left and lower right corners

c and the lower horizontal edge

call upcorner(pij,rij,q,p,nc,isize,2,1,0,2,3,5,6,qmin)
call upcorner(pij,rij,q,p,nc,isize,2,isize-1,1,1,2,4,5,qmin)
call upedgeh(pij,rij,q,p,nc,isize,2,1,2,3,4,5,6,qmin)

write the last line of probabilities to the output file

do 95 i=1,nc
write(3)(p(i,j),j=1,isize)
95 continue

close(3)
end

this is the subroutine to update probabilities at the corner

subroutine upcorner(p,r,q,up,mc,isz,ir,iff,id,k1,k2,k3,k4,qm)
dimension p(3,isz,mc),r(mc,mc,9),q(mc),up(mc,isz)

do 10 i=1,mc
q(i)=0.0
10 continue
do 20 ii=1,mc
do 20 jj=1,mc
q(ii)=q(ii)+r(ii,jj,k1)*p(1,1,f,jj)+r(ii,jj,k2)*p(1,1,f+1,jj) +r(ii,jj,k3)*p(2,1,f,jj)+r(ii,jj,k4)*p(2,1,f+1,jj)
20 continue
a=0.0
do 30 i=1,mc
a=a+p(ir,i,f+id,i)*(mc+1+q(i))
if(q(i).lt.qm)then
qm=q(i)
endif
30 continue
do 40 i=1,mc
up(i,iff+id)=p(ir,i,f+id,i)*(mc+1+q(i))/a
40 continue
end

This is the subroutine to update probabilities at edge-horizontal

subroutine upedgeh(p,r,q,up,mc,isz,ir,k1,k2,k3,k4,k5,k6,qm)
dimension p(3,isz,mc),r(mc,mc,9),q(mc),up(mc,isz)

do 15 j=2,isz-1
do 10 i=1,mc
q(i)=0.0
10 continue
do 20 ii=1,mc
do 20 jj=1,mc
q(ii)=q(ii)+r(ii,jj,k1)*p(1,1,j-1,jj)+r(ii,jj,k2)*p(1,1,j,jj) +r(ii,jj,k3)*p(1,1,j+1,jj)+r(ii,jj,k4)*p(2,1,j-1,jj) +r(ii,jj,k5)*p(2,1,j,jj)+r(ii,jj,k6)*p(2,1,j+1,jj)
20 continue
a=0.0
do 30 i=1,mc
a=a+p(ir,j,i)*(mc+1+q(i))
if(q(i).lt.qm)then
qm=q(i)
endif
30 continue
do 40 i=1,mc
up(i,j)=p(ir,j,i)*(mc+1+q(i))/a
40 continue
15 continue
end

This is the subroutine to update probabilities at edge-vertical

subroutine upedgev(p,r,q,up,mc,isz,ir,k1,k2,k3,k4,k5,k6,qm)
dimension p(3,isz,mc),r(mc,mc,9),q(mc),up(mc,isz)
do 10 i=1,mc
  q(i)=0.0
  continue
  do 20 ii=1,mc
    do 20 jj=1,mc
      q(ii)=q(ii)+r(ii,jj,k1)*p(1,iff,jj)+r(ii,jj,k2)*p(1,iff+1,jj)
      +r(ii,jj,k3)*p(2,iff,jj)+r(ii,jj,k4)*p(2,iff+1,jj)
    2  +r(ii,jj,k5)*p(3,iff,jj)+r(ii,jj,k6)*p(3,iff+1,jj)
    20  continue
    a=0.0
    do 30 i=1,mc
      a=a+p(2,iff+ir,i)*(mc+1+q(i))
      if(q(i).lt.qm)then
        qm=q(i)
      endif
    30  continue
    do 40 i=1,mc
      up(i,iff+ir)=p(2,iff+ir,i)*(mc+1+q(i))/a
    40  continue
  end
REFERENCES


Fox, L., Ill, J. A. Brockhaus, and N. D. Tosta, 1985, Classification of Timberland Productivity in Northwestern California Using Landsat, Topographic, and Ecological


139


Li, B., 1987, *Texture Measurement of Spatial Patterns on Thematic Mapper Imagery: An Experiment*. Master Thesis (Omaha, NE: University of Nebraska at Omaha).


