

2000

Integrating spatial and spectral information for automatic feature identification in high -resolution remotely sensed images

Jong Yeol Lee
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Lee, Jong Yeol, "Integrating spatial and spectral information for automatic feature identification in high -resolution remotely sensed images" (2000). *Graduate Theses, Dissertations, and Problem Reports*. 1212. <https://researchrepository.wvu.edu/etd/1212>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**INTEGRATING SPATIAL AND SPECTRAL INFORMATION
FOR AUTOMATIC FEATURE IDENTIFICATION IN HIGH
RESOLUTION REMOTELY SENSED IMAGES**

Jong Yeol Lee

**Dissertation submitted to the
Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Doctor of Philosophy
in
Geography**

**Timothy A. Warner, Ph.D., Chair
Haluk Cetin, Ph.D.
Gregory A. Elmes, Ph.D.
Trevor M. Harris, Ph.D.
M. Duane Nellis, Ph.D.**

**Department of Geology and Geography
Morgantown, West Virginia
2000**

**Key words: Remote Sensing, Texture, Edge Detection, Image Segmentation,
Multivariate Classification, Object-based Classification
Copyright 2000 Jong Yeol Lee**

ABSTRACT

Integrating Spatial and Spectral Information for Automatic Feature Identification in High Resolution Remotely Sensed Images

Jong Yeol Lee

This research used image objects, instead of pixels, as the basic unit of analysis in high-resolution imagery. Thus, not only spectral radiance and texture were used in the analysis, but also spatial context. Furthermore, the automated identification of attributed objects is potentially useful for integrating remote sensing with a vector-based GIS.

A study area in Morgantown, WV was chosen as a site for the development and testing of automated feature extraction methods with high-resolution data. In the first stage of the analysis, edges were identified using texture. Experiments with simulated data indicated that a linear operator identified curved and sharp edges more accurately than square shaped operators. Areas with edges that formed a closed boundary were used to delineate sub-patches. In the region growing step, the similarities of all adjacent sub-patches were examined using a multivariate Hotelling T^2 test that draws on the classes' covariance matrices. Sub-patches that were not sufficiently dissimilar were merged to form image patches.

Patches were then classified into seven classes: Building, Road, Forest, Lawn, Shadowed Vegetation, Water, and Shadow. Six classification methods were compared: the pixel-based ISODATA and maximum likelihood approaches, field-based ECHO, and region based maximum likelihood using patch means, a divergence index, and patch probability density functions (pdfs). Classification with the divergence index showed the lowest accuracy, a kappa index of 0.254. The highest accuracy, 0.783, was obtained from classification using the patch pdf. This classification also produced a visually pleasing product, with well-delineated objects and without the distracting salt-and-pepper effect of isolated misclassified pixels. The accuracies of classification with patch mean, pixel

based maximum likelihood, ISODATA and ECHO were 0.735, 0.687, 0.610, and 0.605, respectively.

Spatial context was used to generate aggregate land cover information. An Urbanized Rate Index, defined based on the percentage of Building and Road area within a local window, was used to segment the image. Five summary landcover classes were identified from the Urbanized Rate segmentation and the image object classification: High Urbanized Rate and large building sizes, Intermediate Urbanized Rate and intermediate building sizes, Low urbanized rate and small building sizes, Forest, and Water.

Acknowledgement

I am first and foremost grateful for my splendid dissertation advisor Dr. Timothy A. Warner. I am indebted to him in many different ways, including spiritually and financially at the last moment, and this work would not exist without his excellent guidance and considerable encouragement. I will never forget the warm heart of Dr. Warner's wife, Paula Hunt.

I would like to acknowledge all of my committee members: Dr. M. Duane Nellis, Dr. Trevor M. Harris, Dr. Gregory A. Elmes, and Dr. Haluk Cetin. They served on my committee with great understanding, knowledge and patience. I would also like to extend my appreciation to Dr. Kenneth C. Martis, Dr. Robert Q. Hanham, Dr. Daniel Weiner, Dr. Ann M. Oberhauser, and Dr. Calvin O. Masilela. They have been great supporters throughout my time at West Virginia University. Gratitude is also expressed to all my graduate student colleagues, especially Mr. Shawn Banasick. They all helped me overcome this long difficult time.

I am grateful to the Korea Research Institute for Human Settlements for allowing me to conduct my study. The institute supported me financially for such a long time period and generously granted me additional leaves. I would also like to thank all the members of Morgantown Korean Church for their kindness and support.

Finally, I would like to say to my family that I have missed them from the bottom of my heart every moment during my staying at Morgantown.

Table of Contents

ABSTRACT	ii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
TABLE OF FIGURES	viii
TABLE OF TABLES	x
I. INTRODUCTION	1
1 - 1. Research objectives	3
1 - 2. Research hypotheses	5
1 - 3. Description of terminology	7
1 - 4. Organization of the dissertation	9
II. REMOTE SENSING USING HIGH SPATIAL RESOLUTION IMAGERY ...	12
2 - 1. Scene models in remote sensing	12
2 - 2. Spatial resolution and scene objects in remote sensing	14
2 - 3. Incorporating spatial information in image processing	16
2-3-1. Pixel-based approaches	17
2-3-2. Object-based feature extraction	20
2-3-3. Summary	23

2 - 4.	Conceptual Framework	24
2 - 5.	Data and study area	26
III.	IMAGE SEGMENTATION	31
3 - 1.	Introduction	31
3 - 2.	Review of image segmentation literature	32
3 - 2 - 1.	Region-based approaches	32
3 - 2 - 2.	Edge-based approaches	34
3 - 3.	Image segmentation methods	38
3 - 3 - 1.	Image enhancement	40
3 - 3 - 2.	Ridge finding	47
3 - 3 - 3.	Thinning	49
3 - 3 - 4.	Pruning	50
3 - 4.	Application to simulated and real data	51
IV.	REGION GROWING WITH IMAGE SEGMENTS	54
4 - 1.	Introduction	54
4 - 2.	Thresholding	57
4 - 3.	Preliminary extraction of vegetation edges	59
4 - 4.	Sub-patch topology and database development	62
4 - 5.	Region growing with image segments	67
4 - 6.	Application to Morgantown ADAR data	71

V. IMAGE CLASSIFICATION WITH A REGION-BASED APPROACH	75
5 - 1. Introduction	75
5 - 2. Traditional image classification	76
5 - 3. Methods	80
5 - 3 - 1. Classification with divergence index	82
5 - 3 - 2. Maximum likelihood classification using the patch mean	85
5 - 3 - 3. Region based maximum likelihood classification with pdf	86
5 - 4. Application of the pixel and group-based classification methods	88
5 - 5. Evaluation of the classifications	91
VI. OBJECT-BASED LAND COVER CLASSIFICATION	102
VII. CONCLUSIONS	111
APPENDIX	118
REFERENCES	124

Table of Figures

Figure 1-1.	Research Flow Chart.	10
Figure 2-1.	Nested model of urban land use/land cover.	25
Figure 2-2.	Study area of Morgantown.	30
Figure 3-1.	The image segmentation process of identifying sub-patches in the image.	39
Figure 3-2.	Spatial profiles for three directions for a subset of the green band of the Morgantown ADAR image.	42
Figure 3-3.	The eight directions associated with a linear operator five pixels in extent.	43
Figure 3-4.	First order derivatives for three directions for the image shown in Figure 3-2.	43
Figure 3-5.	A comparison of edge enhancement images for three simulated images (A, B, and C) using a directional linear operators and traditional rectangular shaped kernels.	45
Figure 3-6.	Edge enhancement of Morgantown ADAR data for a subset of the study area, including downtown and parts of the neighboring suburbs.	46
Figure 3-7.	A small test subset of the ADAR data of Morgantown, and the results of the ridge finding process.	49
Figure 3-8.	The results of the thinning and pruning process.	50
Figure 3-9.	Comparison of edge detection results for directional linear (rows with upper case letters) and square operators for three sizes of windows/operators.	52
Figure 3-10.	The sub-patch boundaries identified in the initial image segmentation stage of the Morgantown ADAR data overlain on a standard false color composite of the scene.	53
Figure 4-1.	Overview of the region growing stage and patch identification.	56
Figure 4-2.	Comparison of threshold levels of 30, 70, 100, 150 for image sub-patch boundary identification for a small test site within the Morgantown data.	58

Figure 4-3.	Image sub-patches delineated by an edge enhancement threshold level of 70.	60
Figure 4-4.	Preliminary vegetation areas and associated edge pixels.	63
Figure 4-5.	Image segments of sub-patches modified by vegetation edges.	64
Figure 4-6.	An example of the sub-patch segmentation, and the associated polygon identification codes for selected polygons.	66
Figure4-7.	Comparison of different significant levels applied to the region growing procedure.	73
Figure 4-8.	Patches produced by region growing with a significance level of 0.0005.	74
Figure 5-1.	The decision rule of a pixel-based maximum likelihood classifier.	80
Figure 5-2.	Comparison of object-based classification with traditional image classification approaches.	81
Figure 5-3.	The decision rule for $L_j(X)$. a is the probability for class i $p(X w_i)$, and b is the probability for class j $p(X w_j)$	83
Figure 5-4.	Likelihood measured with pdf.	86
Figure 5-5.	Outline of the spectral classification procedure using previously segmented data.	89
Figure 5-6.	Maximum likelihood calculation utilizing patch pdfs.	91
Figure 5-7.	Results of the classifications.	95
Figure 6-1.	Land cover type map classification.	103
Figure 6-2.	Land cover properties of regions.	106
Figure 6-3.	Land cover classification map.	107
Figure 6-4.	The vectorized land cover map.	109

Table of Tables

Table 1-1.	Major terms used in this research.	7
Table 2-1.	Minimum Spatial Resolution Requirements for Land Use/Land Cover Type.	15
Table 2-2.	Spectral properties of the ADAR System 5500 data (Positive Systems, 1997, unpublished data).	26
Table 5-1.	Summary accuracy statistics for seven classes by the six classification methods used in this study.	92
Table 6-1.	Land cover properties by region.	105

Chapter I

INTRODUCTION

The automated mapping of map features such as roads and buildings in remotely sensed imagery can be a highly complex process. This is because automated feature extraction involves the analysis of multidimensional spectral, spatial and contextual relationships, and is analogous to the perceptual and cognitive processes of the human vision system (Hodgson, 1998). It is thought that a fundamental process in human vision is the recognition of more abstract and complex information, particularly edges (Marr, 1982) and multi-scale patterns (Woodcock and Harward, 1992; Ford and McKeown, 1992; Baumgartner *et al.*, 1997). This recognition process may work in two directions. Firstly, local regions may be examined in detail in order to identify fine-scale features, which are then aggregated to objects at coarser scales (Woodcock and Harward, 1992). Alternatively, the process may work in reverse, for example, starting with the recognition of generic objects, and then moving to the identification of attributes and finer scale features (Ford and McKeown, 1992). For this second approach, the human ability to use prior knowledge and heuristic understanding is important in the identification of the

generic objects. Because of the difficulties of duplicating these procedures in computer algorithms, it would appear that the first approach, that of generalizing from fine features to coarser objects, would have more potential for remote sensing-based feature extraction.

One of the reasons for the past difficulties in applying spatial analysis methods to remotely sensed data is the traditional focus on a single element of the data structure, the individual raster pixel. The focus on the pixel has been driven in part by the very successful use of aspatial statistics in image classification. With its 80 meter Instantaneous Field of View (IFOV), the Landsat MultiSpectral Scanner (MSS), the first major civilian earth-observing sensor, had sufficiently coarse resolution that the internal heterogeneity in most land use classes was smoothed, thus reducing the importance of spatial information. The higher spatial resolution of later Thematic Mapper (TM) and Satellite Pour l'Observation de la Terre (SPOT) produced improved delineation of fine detail, but somewhat counter-intuitively produced a lower overall accuracy for land use studies (Cushnie, 1987). This is because the 20-30 meter pixel size of these sensors is close to the scale of the objects within each land use class. Woodcock and Strahler (1987) have shown that under such circumstances a scene has the greatest variability in pixel values, and consequently the lowest potential for conventional classification. This has led to a concentration on texture and other measures of local variability, to improve classification at these scales (Møller-Jensen, 1990; Hay *et al.*, 1996; Ryherd and Woodcock, 1996; Kuroso *et al.*, 1999).

With the advent of high resolution, 1-4 meter data (Kramer, 1996), new approaches to classification will be required. In particular, spatial processing may become even more

important. This may partly be driven by an interest in object identification, instead of land cover mapping. With high resolution data, individual objects such as buildings and roads become potentially resolvable. Remote sensing-based classification could possibly be much more useful if objects, not arbitrary pixels, were the basic structural element of map products. Furthermore, research has suggested that classification can be much more accurate when spatial information is utilized (Kettig and Landgrebe, 1976; Treitz and Howarth, 2000). Perhaps most important of all, the recognition of objects will facilitate the incorporation of remotely sensed data within a GIS (Janssen and Molenaar, 1995; Cleynenbreugel *et al.*, 1990).

1-1. Research objectives

The fundamental goal of this research is to seek computer-based methods to extract spatial information for urban and peri-urban areas from high spatial resolution remotely sensed data. This research has four objectives, which are listed and then discussed in more detail below:

1. Develop a method of segmenting images into discrete image objects.
2. Classify the image objects with spectral and spatial information.
3. Aggregate the extracted information in order to map local land use/cover at an Anderson *et al.* (1976) Level II.
4. Compare the hierarchical image object classification developed in this research to standard aspatial classification methods.

The first objective of this research is to develop a method to segment remotely sensed data as the initial step in automatic feature recognition for classification of urban areas using high resolution imagery (1m or finer). The discrete scene model (Woodcock and Srahler, 1987; Woodcock and Harward, 1992) is employed. The scene is conceptualized as comprising discrete objects, such as buildings, roads, and lawns. The initial segmentation draws on the concept of directional texture (Warner *et al.*, 1999). This segmentation, however, is unlikely to produce groups of pixels that directly correspond to the objects of interest. The main causes of this include the presence of shadow and topographically induced variations in illumination, as well as variations in the spectral properties within objects. Therefore, segmentation is carried out in a hierarchical fashion, first identifying groups of pixels that may be part of an object, defined here as sub-patches. Adjacent sub-patches are then aggregated to form patches, which are in turn aggregated to form objects, for example, houses or roads.

The second objective is to develop a classification method in which image segments (objects) are the basic classification unit (Kettig and Landgrebe, 1976; Bryant, 1990). This approach is in contrast to standard pixel-based classification, in which each pixel value is individually compared with the training area's statistics. Due to internal variation within each object, pixel-based classification tends to have randomly distributed misclassified pixels. This is particularly distracting for visual interpretation, and can cause highly complex vectorization. Classification using image segments overcomes this problem because the variance, and potentially the covariance, of image segments can be used for classification.

The third objective is to produce a thematic map by aggregating the classified objects to the next hierarchical layer. The identification of low level scene objects can potentially result in a much greater accuracy in mapping at the next hierarchical level. This is because the land cover, and by inference land use, is mapped from the composition and proportion of scene objects such as buildings and roads , and not by the average spectral characteristics of the land use itself.

To accomplish the final objective, that of providing an evaluation of the success of this research, the classification is tested on high spatial resolution data of Morgantown and compared to a conventional aspatial classification (Richards, 1994) and the ECHO (Extraction and Classification of Homogeneous Objects) field-based classifier (Kettig and Landgrebe, 1976). The ground reference data for evaluation is derived from a combination of expert visual interpretation of the imagery and local knowledge.

1-2. Research hypotheses

The research objectives outlined above are based on several hypotheses regarding ways of incorporating spatial properties within an image analysis procedure. These hypotheses, as well as the assumptions on which they are based, are discussed in more detail in this section.

Firstly, it is assumed that with high-resolution data the spectral characteristics of individual objects, such as individual buildings and roads, are relatively homogeneous, and that objects are distinguishable against the background and adjacent objects. Therefore, the edges of individual objects will be characterized by a relatively high

spectral gradient, manifested as a "ridge" of high local variance surrounding the lower values associated with the interior of objects. Theoretically, this ridge should form closed loops that delineate objects. This leads to the first hypothesis, that edge detection using a "ridge finding" procedure applied to local variance will be an effective method of image segmentation.

The second hypothesis is that the location of edges will be detected more accurately by multiple, centrally co-incident, linear shaped edge detection operators, rather than traditional square shaped windows. Because of the differences between the geometric characteristics of edges and square shaped windows, the precision of the edges detected will be lower with square windows. For example, with a square window, the derived edges on a sharp corner of an object will not follow the object edge, but will rather be rounded. This will be most serious for large window sizes. This is significant, because large windows are generally needed for high spatial resolution images (Ferro, 1998). The data for this experiment comprise simulated images of geometric objects on a background. The results for edge detection using linear operators are compared to those obtained using square windows.

The final hypothesis is that a region-based classification will result in a higher classification accuracy than traditional pixel-based classification for high spatial resolution imagery. The variability of spectral characteristics of individual pixels within an object is likely to increase for high spatial resolution imagery. Consequently, this will reduce the efficacy of pixel-based classification approaches. By contrast, for a region-based approach, the spectral variability can be a part of the characteristics used for classifying a region. The data used to test this hypothesis is aircraft-acquired digital

imagery of Morgantown, WV. Traditional classification, ECHO (Kettig and Landgrebe, 1976), and the method developed in this research is compared to ground data through the use of standard error matrices.

1-3. Description of terminology

In the previous sections a number of terms have been used in a relatively loose sense. To facilitate a more in-depth discussion, it is important to define the terminology used more rigorously (Table 1-1). The term ‘object’ will be used in this thesis to refer to

Table 1-1. Major terms used in this research

Class	Associated Boundary	Description of Class
Feature	-	Generic object class
Object	Object edge	Entity instance, produced by aggregation of like-classified patches
Patch	Patch edge	Group of pixels of similar radiance, produced by region-growing applied to sub-patches
Sub-patch	Sub-patch edge	Closed polygons formed by texture ridges after branch texture ridges removed
-	Texture ridge	Local high in texture image in at least one direction

the concept of an ‘entity instance’ as defined by the American National Standard for Information Systems (USGS, 1997), and the term ‘feature’ will be used as a generic concept for object classes. Objects are the fundamental units of interest in this study. Variations in illumination caused by differences in slope and aspect, as well as real differences in spectral properties within an object, will cause the spectral segmentation to

subdivide objects into segments smaller than objects. Thus many real objects will be represented in the image by a number of discrete, spectrally separable areas, defined here as 'patches.' A patch itself might also be segmented into spurious subdivisions, termed 'sub-patches', due to the presence of edges internal to a patch such as walls or other two-dimensional features. The sub-patches belonging to a single patch are, however, spectrally similar to one-another, and therefore can be aggregated using multivariate measures of spectral similarity.

The edges are derived from a texture image (Table 1-1). The term 'texture ridge', or more briefly, 'ridge,' is used to refer to the primitive edge pixels that are local highs in spatial profiles drawn in at least one direction in the texture image. The ridges include 'branch ridges,' which are edges within sub-patches that do not delineate groups of pixels (i.e. they do not form closed polygons). Once a ridge has been determined to define a closed polygon, it is re-classed as a sub-patch edge. The boundary pixels of patches are similarly referred to as 'patch edges.' The term 'object edge' is used for the boundary pixels of objects.

The purpose of automated feature extraction is the extraction and identification of image objects, and not necessarily the building of an object-oriented database, although the product of a feature extraction algorithm could be used to support an entity-based model (Worboys, 1994). An advantage of an object identification approach is that it facilitates the integration of remote sensing and GIS (Gahegan and Flack, 1999; Tang *et al.*, 1996). This is because individual objects are more appropriate for incorporation in a GIS environment, especially one that is vector-based (Usery, 1996; Usery, 1993).

Conventional remote sensing classification produces labeled pixels, rather than delineated objects.

A fundamental assumption of this research is that the objects of interest, such as buildings and roads, are clearly resolved by the high spatial resolution imagery. Strahler *et al.* (1986) have labeled such imagery as H- resolution imagery, in contrast to L- resolution imagery in which the pixel size is too large to resolve the individual objects of interest. However, because this research is focused on meter imagery, a scale commonly referred to as “high resolution,” this term will be used in preference to the term H- resolution imagery.

Several terms have been used for the process of rescaling data, including degradation (Pax-Lenny and Woodcock, 1997), upscaling (Hay *et al.*, 1997), aggregation (Kloditz *et al.*, 1998; Bian and Butler, 1999), as well as upward and downward resampling. In this study, the term 'aggregation' will be used for resampling imagery to lower resolutions (Kloditz *et al.*, 1998; Bian and Butler, 1999).

1-4. Organization of the dissertation

This dissertation has seven chapters, including this chapter. Following the introductory material in Chapter I, Chapter II summarizes the theoretical background and previous research. Scene models in remote sensing, the relationship between object identification and spatial resolution, the exploitation of spatial information in remote sensing, and previous approaches to image segmentation for feature extraction are discussed.

The following four chapters discuss the four steps of the image analysis procedure developed in this research (Figure 1-1). In each chapter additional literature is presented, and both methods and results are discussed. Two types of data are used to develop and test the methods investigated in this research. Simulated data is used where it is important to have absolute knowledge of the data and the scene it represents. Aircraft imagery of Morgantown, West Virginia is used to test real world applications of this research.

Chapter III describes the proposed ridge following edge detection method, and compares image segmentation based on this method to the results obtained from a square texture window. Chapter IV is an investigation of image segmentation based on the amalgamation of patches identified through ridge-following edge detection. A central aspect of this work is the development of a topology for the patches to facilitate the analysis of the spatial relationships between adjacent patches.

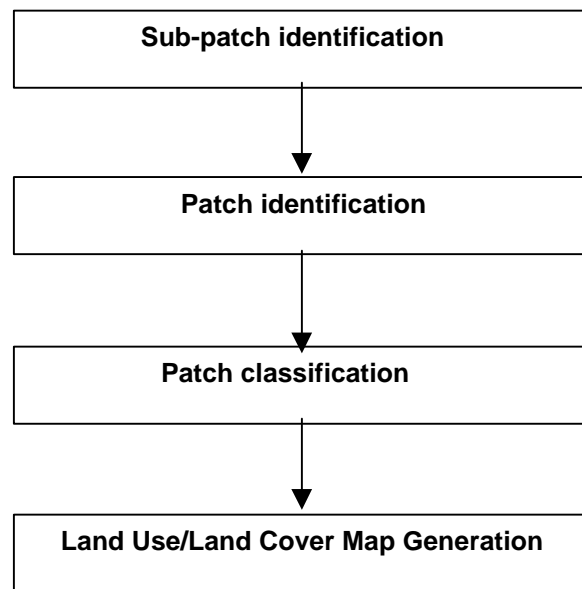


Figure 1-1. Research Flow Chart

Chapter V proposes a region-based classification scheme using the image elements derived in the image segmentation. Three approaches for the classification using patch statistics are compared: classification using the patch mean, a divergence index, and patch probability density functions (pdfs). The results from these methods are compared with those derived from various traditional and field-based classifications. Chapter VI discusses feature-based land cover classification. A classification is produced from the proportion of the image objects within each class, instead of the average spectral properties of the class. This method makes it possible to classify imagery hierarchically. For example, density of residential area is produced based on the percentage of built-up area and density of building polygons per unit area. Finally, Chapter VII gives the overall conclusions and suggestions for future studies of high spatial resolution remotely sensed data.

Chapter II

REMOTE SENSING USING HIGH SPATIAL RESOLUTION IMAGERY

This work is designed to exploit the current generation of high resolution panchromatic and multispectral sensors (Kramer, 1996). The research draws on a theoretical understanding of the spatial properties of high resolution data to develop new methods appropriate for the scale of such imagery. This section, therefore, places the research in its theoretical context, demonstrating its conceptual heritage and the significance of the research problem.

2-1. Scene models in remote sensing

Scene models in remote sensing originate from conceptual developments in the understanding of imagery. There are two main scene models in remote sensing: the discrete model and the continuous model (Strahler *et al.*, 1986). For the discrete model, the scene is conceptualized as discrete objects on a background. The objects have

distinctive, homogeneous spectral properties, which potentially allow identification of the class to which the object belongs. Therefore, an instantaneous field of view (IFOV) smaller than the object is important in order to ensure that the distinctive spectral properties of the object can be identified.

The discrete model is separated into three sub-models based on the scene composition (Strahler *et al.*, 1986). Simple models comprise scenes that consist of a single object and background, whereas complex models deal with scenes that are composed of multiple classes of objects or backgrounds. Nested models are multi scale (Woodcock and Harward, 1992), with each level in the hierarchy comprised of aggregations of objects or classes from the next finer level of objects or classes (Woodcock and Strahler, 1987). For example, in the Anderson *et al.* (1976) scheme, the Urban and Built-up Land Level I class is composed of seven level II classes, such as Residential and Industrial. Finer levels of detail, for example Buildings, Roads, Lawns and Trees, are specified on a project-specific basis.

The continuous model assumes that the changes in the scene are continuous, not abrupt. Consequently, in an image of such a scene, continuously changing spectral properties are portrayed. Individual objects cannot be resolved (Collins and Woodcock, 1999a; Foody, 1999). An example of a real world analogue of the continuous model is a field with varying densities of crops, imaged with a sensor that has an IFOV larger than the size of the individual plants. The main concern of the continuous model is the relationship between the IFOV and the spectral variability in the scene (Collins and Woodcock, 1999a; Henebry, 1993). However, even with the continuous model, regions with certain average or statistical properties can usually be identified. If so, the result of

an analysis can sometimes take the form used for the discrete model. Otherwise, a fuzzy approach may be appropriate for analysis of continuous models (Bezdek, 1993; Warner and Shank, 1997a; Foody, 1999).

This study assumes that individual buildings, roads, and patches of vegetation may be discerned with high spatial resolution remotely sensed data, and can be represented in corresponding detail in the land use/land cover maps derived from such data. This detail can be aggregated to coarser scales for map display. Therefore, this research is embedded within the conceptual framework of the nested discrete model. However, the edge detection methods used in this research focus on mixed pixels on the boundaries between cover types, and therefore in a sense draw on a continuous model of the scene.

2-2. Spatial resolution and scene objects in remote sensing

The spectral variability of remotely sensed data is determined by the spatial properties of the sensor (Collins and Woodcock, 1999b), as well as the spatial properties of the spectral classes (Fung and Chan, 1994; Marceau *et al.*, 1994a), as represented by the size and arrangement of the objects making up the classes (Woodcock *et al.*, 1988a; Woodcock *et al.*, 1988b). The hierarchical nature of these scene objects is the basis for the USGS land-use/land-cover classification system (Anderson *et al.*, 1976; Woodcock and Harward, 1992). Because of the complex nature of remotely sensed images, the finer the detail that is observed, the lower the hierarchical level of the classes that can be mapped. Jensen and Cowen (1999) reported the relationship between hierarchical level of classes and the corresponding spatial resolution required (Table 2-1). For example, mapping USGS land-use/land-cover level I classes requires 20-100 meter resolution,

whereas delineating building perimeters requires a spatial resolution of 0.25-0.5 meter. To identify individual objects of interest, an IFOV that is at least several times smaller than objects is required (Welch, 1982).

One of the most important studies to illustrate the issue of image scale and its potential influence on classification accuracy, is that by Woodcock and Strahler (1987). They found that the local variance, when plotted as a function of spatial resolution, reached a peak at a spatial resolution 50 – 75 percent of the average size of objects in

Table 2-1. Minimum Spatial Resolution Requirements for Land Use/Land Cover Type. (After Jensen and Cowen, 1999.)

Land Use/Land Cover Type	Minimum Spatial Resolution Requirements
USGS Land Use/Land Cover Level I	20-100 meters
USGS Land Use/Land Cover Level II	5-20 meters
USGS Land Use/Land Cover Level III	1-5 meters
USGS Land Use/Land Cover Level V	0.25-1 meter
Building Perimeter	0.25-0.5 meter

the scene. This scale produces the greatest proportion of mixed pixels, and is the least desirable for automated classification because of the high degree of variance that does not contribute to spectral discrimination of the classes of interest (Latty *et al.*, 1985). Atkinson and Curran (1997) developed this concept further, using a semivariogram to quantify image scale. Warner and Shank (1997b) suggest that the optimum spatial resolution could be defined as the resolution that clearly resolves the objects of interest, but maximizes the spectral mixing of objects and the next finer hierarchical level. The

optimal scale is thus one at which both the within-class and between class variance is minimized.

The previously mentioned studies have focused on identifying a single, most appropriate spatial resolution for an image. However, the optimal spatial resolution is likely to vary with the size of the objects in an image. Marceau *et al.* (1994a; 1994b) evaluated the impact of spatial resolution and spatial aggregation on the classification accuracy of very high spatial resolution (3.7-m to 0.5-m) images of tree stands. They found that coarsening the spatial resolution reduced image variance, but the effects of aggregation were different for different objects and spatial structures. Hay *et al.* (1997), using images with pixel sizes of 1.5, 3, 5, and 10-m, showed that each of the six forest classes studied has its own most appropriate scale. Marceau *et al.* (1994a; 1994b) and Hay *et al.* (1997) argued against the concept of a single optimal scale, and noted that no single scale is capable of accurately describing the various sizes and shapes of objects within a remotely sensed scene. These studies therefore suggest that the size and spectral properties of objects under investigation should be defined clearly in order to determine the most appropriate scales of analysis (Marceau *et al.*, 1994b).

2-3. Incorporating spatial information in image processing

Traditional classification methods in remote sensing were designed to deal with moderate resolution satellite images that have 30-80 meter pixels (Swain and Davis, 1978). These methods typically employ an aspatial, pixel-based approach for image classification that is not very effective for classifying high spatial resolution data

(Gougeon, 1995a). Spatial analysis of imagery remains one of the most important areas of research in remote sensing, because of its potential for increasing classification accuracy, especially for high spatial resolution data. The major methods that have been investigated include reduction of noise or within class variance through the aggregation of pixels and low pass filtering, texture analysis, and image segmentation. Generally, these methods are focused on modifying, or converting the spatial measurements into pixel data that can be regarded as an additional spectral band for the purposes of classification. Therefore, these methods are referred as pixel-based approaches.

2-3-1. Pixel based approaches

There are many empirical investigations of aggregating spatial resolution to find the most appropriate spatial resolution for particular applications and to quantify the effect of spatial resolution on classification accuracy (Pax-Lenney and Woodcock 1997; Teillet *et al.*, 1997; Latty *et al.*, 1985). Pax-Lenney and Woodcock (1997) assessed the effect of spatial resolution on change detection accuracy for agricultural land-use in the Western Desert adjacent to the Nile delta. The accuracy of change detection products using TM data and four lower resolution data sets was found to decrease slightly as spatial resolution decreased. However, they suggested that the coarse spatial resolution images were valuable because of the smaller data volume. Teillet *et al.* (1997) produced a similar analysis for forested regions, and concluded that aggregations in spatial resolution were influenced by specific land cover characteristics. Latty *et al.* (1985) also performed an analysis to test the classification accuracy according to the different spatial resolution

sizes for Landsat TM data. TM data were aggregated from 28.5 to 57 meter resolution, and the data sets were classified with pixel-based and field-based methods. The mean accuracy observed for the 57 meter aggregated data was higher than that obtained from the finer scale original data, except when the field-based ECHO (Kettig and Landgrebe, 1976) classifier was used.

Although aggregation can be useful for selecting optimal scales, aggregation inherently implies a loss of the shape information that may be useful in high resolution image analysis. Spatial convolution filtering, used to suppress the high spatial frequencies in imagery (Jensen, 1996), may be an alternative to aggregation. In this method, low pass filters are used to generate pixel values representing the average of a local region of pixels (Meyer *et al.*, 1996; Hay *et al.*, 1996). A problem with all convolution methods is that the effect of spectral mixing between classes, sometimes referred to as the edge effect (Ryherd and Woodcock, 1996), tends to dominate over the information of interest, the within class variance, and may result in an unacceptably poor classification.

The methods discussed so far focus on modifying spectral radiance measurements to suppress unwanted variation. In contrast to this approach, many methods have been developed to quantify the local spectral variability. One of the most common uses local properties obtained from a moving window, especially the statistical properties of variance and standard deviation (Woodcock and Strahler, 1987). It is interesting to note that with this approach, the spatial arrangement of the pixel values within the window itself is not measured in any way.

The gray level co-occurrence matrix (GLCM) provides conditional probabilities of a matrix based on the frequency of the co-occurrence for all pairs of spectral values at a certain distance and angle in a window (Haralick *et al.*, 1973). It is therefore unlike the variance-based filter, in that it incorporates measures of the spatial arrangements of Digital Numbers (DN values) within the window. Examples of the spatial arrangements that are measured with the GLCM include uniformity, correlation, and entropy. Barber and LeDrew (1991) demonstrated the discrimination power of GLCM.

Wang and He (1990) developed a texture measurement based on comparing the spectral magnitude between various orderings of pixels within a window. For example, a 3 by 3 window has 256 possible combinations of spatial orderings. The different combinations of the data within the window are counted, and a histogram of data combinations is produced. Three texture measurements are defined based on the histogram, including black-white symmetry (BWS), geometric symmetry (GS), and degree of direction (DD). Gong *et al.* (1992) used these texture measurements to compare spatial feature extraction algorithms with SPOT HRV data. However, these spatial features added only limited improvements to the overall classification accuracy that could be achieved using the multispectral data alone.

The semivariogram was developed to represent the spatial dependence between each point and its neighbors over a range of distances, known as lags (Matheron, 1963; Journel and Huijbregts, 1978; Curran, 1988). The average variance of pixels at various lagged distances generally increases as the lag increases for short ranges, but reaches a sill at a lag where the pixels are no longer correlated with their neighbors (Curran, 1988). St-

Onge and Cavayas (1997) have shown the value of the semivariogram for mapping forest structure from high spatial resolution imagery.

Image segmentation is another widely used method of reducing spatial variability prior to classification. For example, local variance was used for segmenting of forest images by Ryherd and Woodcock (1996), and Hay *et al.* (1996). Following segmentation, summary spectral information on pixel clusters, for example the mean spectral measurement (Gougeon, 1995a; Hay *et al.*, 1996; Meyer *et al.*, 1996), is used to characterize the clusters as a single group. Thus, the spatial variability within the cluster is often ignored. Euclidean spectral distance between cluster means has generally been used as a criterion to merge the groups of adjacent pixels in image segmentation methods (Woodcock and Harward, 1992; Franklin and Wilson, 1991), and for classification.

The spatial structure of the tonal values is stressed in texture transformations, and the spectral characteristics of the objects are ignored. In this respect, spectral and spatial approaches are mutually complementary and have been combined in some studies (for example, Ryherd and Woodcock, 1996; Meyer *et al.*, 1996; Hay *et al.*, 1996).

2-3-2. Object-based feature extraction

With high resolution data, it should be possible to map individual objects, such as buildings in urban areas, for example. Thus classification methods could evolve from an approach based on the spectral characteristics of individual pixels in classes, to an object-based approach. A study that gives support to this idea is that of Hodgson (1998), who examined the way humans identify urban classes. He carried out experiments using 1.5

meter data and image subsets of a range of sizes. He found that the smaller the window size, the more the classification depends on the spectral characteristics of objects. However, with increasing window sizes, the human subjects were progressively able to incorporate the context of the scene. For example, a minimum area of 40 by 40 pixels was needed for classification of Anderson level II land use. It can therefore be postulated that for classification human vision requires information not only on the individual objects of interest in a scene but also the context of those objects. This suggests that an object-based nested model might be an effective method of automated classification of remotely sensed imagery. Such an approach could be employed to convert pixel brightness values into spatial information.

The first stage of an object-based approach is to identify segments or groups of pixels that may correspond to objects or parts of objects. For example, in his tree identification studies, Gougeon (1992; 1995a) separates forest area and background based on the spectral properties of these classes, and identifies individual tree crowns using a valley following method that exploits the relatively low spectral radiance associated with the shadowed regions between trees in coniferous forests. More than 80 percent of individual trees were successfully identified in a mixed red pine and white spruce forest using this scheme. In a forest study using coarser resolution data, Woodcock and Harward (1992) identified image segments using a region growing method based on the spectral similarity of neighboring pixels. Trinder *et al.* (1997) found segments of roads using an edge detection method. The segments are combined into a road network by exploiting spatial knowledge concerning the characteristics of roads, for example the association of sub-parallel line pairs.

In most segmentation studies, the segments are classified as a unit based on summary spectral properties. On the other hand, Meyer *et al.* (1996) first classified the individual pixels with maximum likelihood and parallelepiped techniques using spectral and textural information, and then used manually digitized crown boundaries to identify a single dominant class for each crown. Janssen and Molenaar (1995) and Gougeon (1995a) also investigated post classification as a method of classifying segments. Post-classification approaches are, of course, of limited value if the purpose of the segmentation is to allow the classification to be carried out over regions, instead of pixels.

The geometry of features in high spatial resolution data has also been used for identifying objects. For example, Sahar and Krupnik (1999) extracted building outlines by first extracting edges from the image, including both relevant building edges and spurious edges. Geometric properties such as parallel lines and corners connecting parallel lines were used to identify corners of buildings, which were then connected to determine building shape. McKeown (1990) identified linear and sharply curving boundaries using an edge detection method. A 'collinear' line linking process is used to combine edges that share the same structure (Shufelt, 1999). One criticism of these methods is that the geometric approach is effective for rectangular or square shape buildings, but may not be effective for irregular shaped objects and complex road networks. Although object-based approaches have advantages over pixel-based approaches for the analysis of high spatial resolution data, object-based approaches generally deal with only simple models. In particular, it is difficult to characterize the rules for complex urban scenes that contain a variety of different objects.

Contextual information can be used to overcome the limitations of simplistic geometric approaches. For example, Baumgartner *et al.* (1997) used contextual radiometric and geometric information to extract a road network using edge detection data from 0.25 meter, high spatial resolution imagery. The road segments separated by shadows cast by high buildings and trees were connected by spatial context rules. For example, if two approximately co-linear road segments were interrupted by a low DN region, the dark region was examined to identify whether it could be a shadow from an adjacent object. If the region was confirmed as a shadow, the separate parts were connected.

2-3-3. Summary

The level of detail provided by new meter and sub-meter scale imagery may support an analysis that is grounded on the discrete scene model in which individual objects are discerned. With high resolution imagery, the use of traditional approaches might result in excessive within-class variation. Traditional approaches addressed such problems by the aggregation of pixels to minimize the within-class variation. However, the potential of high resolution data may not be fully realized with these approaches because they do not support the recognition of individual features and tend to blur shape information.

This dissertation does not focus on thematic approaches, which are the foundation of traditional image classification, but instead will focus on feature extraction using object identification. Edges of the individual objects will be associated with zones of high variance. Thus texture information will be used to delineate groups of homogeneous

pixels with a low within class variance. It will be necessary to extend simple model oriented feature-based approaches to a complex or nested model applicable to more general urban scenes that have various objects. This study will explore the feature-based approach based on a nested model for an urban scene. It is anticipated that such an approach will have the greatest potential in realizing the possibilities of high resolution imagery. In particular, this approach can contribute to the integration of raster and vector data models.

In the next section the conceptual framework, and the data used in the research, will be presented.

2-4. Conceptual Framework

In this research, the concept of a scene will be based on the nested, hierarchical model (Woodcock and Harward, 1992). It is assumed that the scene is composed of patches, which are regions of relatively uniform spectral properties that differ from neighboring patches (Table 2-1 and Figure 2-1). Patches may be sub-divided into sub-patches by the presence of internal boundaries such as walls or narrow shadows, but they have similar spectral properties to adjacent sub-patches of the same spectral class. Patches represent variation at a scale finer than the objects of interest, and are resolved because a fine scale of data capture is required to identify object shapes. Furthermore, the hierarchy of scales is overly simplistic; the distinctive scales of organization actually overlap. Thus to identify the spectral properties of small buildings, a small IFOV is

required. This IFOV may result in the discrimination of unwanted detail in large buildings.

Land cover objects (Table 2-1 and Figure 2-1), which are the fundamental units of interest in this study, include buildings, roads and vegetation, and are derived by merging patches of different spectral classes, but of the same informational class. Land cover objects have a hierarchical structure in the composition of the scene. For example, Residential and Commercial land use classes, although an abstraction from land cover, can usually be inferred from characteristic numbers, sizes and arrangements of land cover units such as buildings, parking lots, etc. This is because human activity and associated economic functions normally impose distinctive local spatial patterns of

Level of abstraction

Regional land use

Land use

Land cover object

Patch

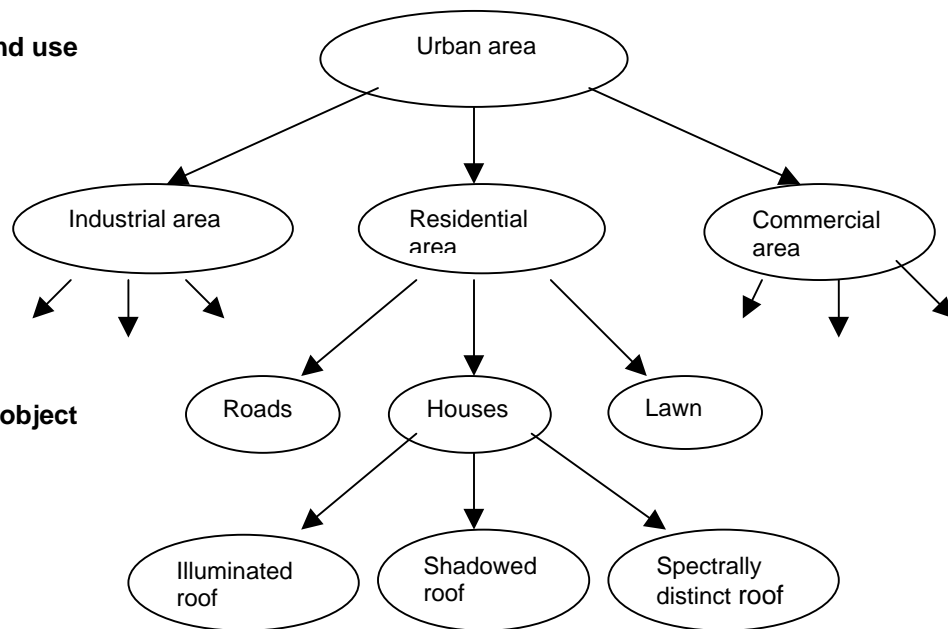


Figure 2-1. Nested model of urban land use/land cover

organization (Lillesand and Kieffer, 1994). There are multiple hierarchical land use classes. For example, Residential and Commercial land use can in turn be aggregated at the next higher hierarchical level to a regional land use class such as Urban.

2-5. Data and study area

The identification of individual buildings and roads requires imagery with a spatial resolution that is much smaller than objects in the scene. For this reason, the spatial resolution of image data should be very high, one meter or less. Because, at the time this project was started, high resolution satellite data was not yet available, Airborne Data

Table 2-2. Spectral properties of the ADAR System 5500 data (Positive Systems, 1997, unpublished data.)

Band	Color	Band Width (nm)	Aperture	Exposure (seconds)
1	Blue	450 – 540	f 2.8	1/500
2	Green	520 – 600	f 2.8	1/800
3	Red	610 – 690	f 4.0	1/640
4	Infrared	780 – 1000	f 5.6	1/2500

Acquisition and Registration (ADAR) 5500 data (Stow *et al.*, 1996) was selected for the analysis. The system captures images 1,000 by 1,500 pixels in size, producing a field of view of 1,000 by 1,500 meters at the nominal IFOV of one meter. The ADAR system acquires four bands of data with four separate digital cameras sensitive to blue, green, red, and infrared wavelengths covering the range from 400 to 1,000 nm (Table 2-2). The separate image bands were co-registered by Positive Systems following the image acquisition.

The data were acquired at 19:42:18 GMT (2:42 pm local time) on March 24 1997 from an altitude of 2,522 meters. Late March in West Virginia is early spring, prior to leaf-out of the deciduous trees. Imagery without the obscuring presence of deciduous leaves minimizes the occlusion of ground objects, and is therefore preferable to summer imagery. In addition, atmospheric visibility tends to be high at that time of year. On the other hand in summer imagery, forest cover would be more spectrally distinctive, and thus easier to identify than in spring imagery.

The study area comprises part of West Virginia University and downtown Morgantown, as well as the adjacent South Park and Woodburn residential areas (Figure 2-1). The scene includes the Monongahela River and its tributary, Decker's Creek. The latter river is bordered by a wooded area.

This area has diverse topographic characteristics. Steep slopes are common adjacent to the rivers. The Downtown area, located between the Monongahela River and Decker's Creek, is relatively flat, with an overall slope to the southwest. The complex topography results in both straight and twisted roads, resulting in a varied road pattern.

A range of land use/land cover classes occur in the study area, including a Central Business District (CBD), various densities of residential housing, institutional facilities, a forested area, lawn, and water bodies. The CBD is composed of relatively tall buildings up to 10 stories high, although the majority are 3-4 stories. The institutional buildings of West Virginia University tend to be larger in size than the other buildings in the area. The residential area to the east and northeast of the study site is hilly, and the houses are located on relatively large plots. These houses are detached, and separated from their

neighbors by lawn and trees. The residential area south of Decker's Creek is relatively dense, with much smaller plots.

Lawn is associated with both public buildings and houses, but not the CBD buildings. Although in West Virginia forests leaf out after late March, when the image was acquired, lawns are typically green. Deciduous forest occurs on both sides of Decker's Creek. The Monongahela River has relatively deep water, and consequently a relatively low overall reflectance. By contrast, Decker's Creek is shallow, and has a higher radiance due to reflectance from exposed and shallowly covered rocks in the streambed.

Different scenes might result in slightly different or additional problems. Rural scenes may have relatively simple land cover classes compared to the selected study area. Furthermore, shadows obscuring neighboring buildings are less likely in rural scenes, although tall structures, such as grain silos and large trees, may cause additional problems. Reduced shadow occlusion should result in a clearer delineation of land cover classes. However, the main determinant of the success of both the image segmentation and classification is likely to be the spectral contrast between the objects of interest. Spectral contrast between objects in rural area, for example, may be more dependent on the seasons than in urban areas because of the smaller proportion of the scene likely to be covered by non-vegetated surfaces.

Urban scenes with taller buildings are likely to suffer from even more shadowing problems than the Morgantown study area. Shadowing may be reduced by careful selection of data acquisition time and season, in order to ensure a high solar angle. Additionally, parallax distortion may cause high buildings to obscure streets and

neighboring buildings, although this can be minimized by acquiring the images from a higher altitude, with a longer lens. In many cases, however, the user has little choice over the acquisition parameters of the imagery used in a study because imagery is often collected for multiple purposes, or previously acquired archived data is used.

Areas with simpler street patterns may be amenable to straightforward geometric rules for identifying the road network. However, an approach based on simple geometry is unlikely to be sufficiently general to be applied across a broad range of cultural and topographic landscapes.

In summary, the study area includes a wide range of topographic settings and land use/land cover types, with a complex road pattern. Classification of such a diverse area presents a challenging task, and a successful approach is likely to be useful over a wide range of regions. Consideration of alternative location may produce different results, though it is very possible that the results could be better than those achieved for the relatively complex urban form of Morgantown.



Figure 2-2. Study area of Morgantown. The study area is approximately 1 by 1.5 kilometers in size. North is towards the top of the image.

Chapter III

IMAGE SEGMENTATION

3-1. Introduction

Image segmentation is a common initial step in object identification from remotely sensed imagery (Barzohar *et al.*, 1997; Ford and McKeown, 1992). The major objective of image segmentation is to partition an image into regions that are spectrally and texturally homogeneous, and that are assumed to delineate the objects or regions in the scene. This approach reduces the complexity of an image, and the delineated objects can be used as a single group in further analysis (Woodcock and Harward, 1992; Coleman and Andrews, 1979; Canny, 1987; Fjørtoft *et al.*, 1998).

There are two major approaches to the segmentation of imagery: region-based and edge-based approaches (Fjørtoft *et al.*, 1998; Haralick and Shapiro, 1985). The region-based approach defines the segments of an image by direct identification of areas with similar spectral or textural values. In contrast, the edge-based approach focuses on the

identification of boundaries through an analysis of tonal or textural change gradients (Fjørtoft *et al.*, 1998).

3-2. Review of image segmentation literature

3-2-1. Region-based approaches

One approach to region-based segmentation is to use **thresholding** of image digital numbers (DN) (Haralick and Shapiro, 1985). Thresholding exploits natural breaks in the frequency histogram of an image, such as local frequency minima, to identify spectral clusters. Groups of spatially adjacent pixels that fall in the same spectral cluster define the individual segments. One-dimensional thresholding uses the histogram for a single gray level image (Haralick and Shapiro, 1985). For multispectral images, thresholding requires multivariate identification of frequency clusters (Coleman and Andrews, 1979).

Although thresholding is simple to implement, it has a number of drawbacks. The initial clustering is aspatial, and therefore there is no *a priori* control over the size or distribution of the spatial clusters. Another problem is that the frequency histogram is developed from global statistics, and this may hinder the identification of overlapping spectral clusters, or the classes that comprise a very small area.

Region growing methods provide an inherently spatial approach compared to the aspatial thresholding methods described above. Region growing is based on annexation processes in which pixels, or groups of pixels, are merged with adjacent pixels or groups that have similar properties. The measures of similarity are typically based on the

statistics of the regions that are being compared, for example, the minimum distance between the region means, and student's *t*-test (Woodcock and Harward, 1992; Franklin and Wilson, 1991). The annexation process is sometimes applied in multiple iterations, so that groups that were formed in one pass through the image may be merged in a later pass.

An important difference between region growing approaches is the spatial unit used in the aggregation process. Single-linkage algorithms compare individual adjacent pixels only (Haralick and Shapiro, 1985; Ryherd and Woodcock, 1996). Single-linkage algorithms tend to produce many small regions because, in the context of graph theory, each pixel is regarded as a node.

The hybrid single-linkage algorithm applies the local statistics of $k \times k$ neighborhood pixels to the center pixel (Haralick and Shapiro, 1985). Conceptually, this approach has similarities to low pass filters and texture measures. Because a larger group of pixels is evaluated for similarity, this method works better on noisy data than single linkage algorithms. However, two potential problems are associated with hybrid single-linkage methods. Firstly, the user has to specify a fixed neighborhood size. Secondly, the neighborhood operator may cause extraneous segmentations near class boundaries.

The field-based approach is particularly suited to multivariate data. In the ECHO classification of Kettig and Landgrebe (1976), homogeneous regions are identified based on the means of fields of multiple univariate data. The initial spatial units are groups of four adjacent pixels. The candidate group's mean for each band is examined sequentially, to determine if the pixels are sufficiently similar to belong to the same region. After the procedure has been applied to the whole image, it is iteratively

reapplied to allow additional merging of the regions. Pixels that are not merged into regions are classified individually.

In contrast to the aggregation approach of region growing image segmentation, **split and merge** methods are a top-down approach (Haralick and Shapiro, 1985). In the first step, the entire image is subdivided into a small number of arbitrary segments. Dividing the image into fourths clearly lends itself to a quadtree approach (Franklin and Wilson, 1991). Each segment produced by the subdivision of the image is tested for internal homogeneity. Regions that meet the homogeneity criterion, for example a student's t -test, are not divided further. Regions that fail the homogeneity tests are recursively divided until all individual regions are considered homogeneous. Subdivided regions can also be aggregated if they meet a similarity criterion (Muerle and Allen, 1968). A limitation of this approach is that the regions are by definition based on geometric shapes, typically rectangles. Furthermore, the results will depend on the user-chosen values selected for the thresholds of internal homogeneity and similarity between regions that determine when to split and merge the segments.

3-2-2. Edge-based approaches

Region-based approaches suffer from an inherent contradiction. Although the purpose of segmentation is usually to identify regions that can subsequently be classified as a group, a preliminary classification is necessary to identify the group itself. Therefore, an alternative is to use edge enhancement to identify areas of significant spectral change that may represent the boundaries of regions. Typically, a threshold is

applied to the edge enhancement output, in order to identify edges. The underlying assumption of this approach is based on the discrete scene model. Real objects or regions in the scene are assumed to be characterized by a homogeneous spectral radiance that contrasts with that of the background across sharp boundaries (Gougeon, 1995b).

Edge detectors, as the name implies, are designed to find the break points between regions with abruptly different spectral radiance (Elder and Zucker, 1998; Fjørtoft *et al.*, 1998). Most edge detectors are based on a gradient analysis, although texture measurements, such as local variance or GLCM indices (Holyer and Peckinpaugh, 1989), are also frequently used. The intensity gradient within the window is defined by:

$$f'(x) = \lim_{x \rightarrow N} \frac{f(x) - f(N)}{x - N}$$

where:

$f(x)$: Spectral radiance of a pixel adjacent to the pixel under investigation

$f(N)$: Spectral radiance of the pixel under investigation

N : Location of the pixel under investigation

x : Adjacent location to N

Unfortunately, this formula cannot be solved directly in the spatial surface of the image. Therefore, the gradient is calculated over a local proxy, the kernel. A number of kernels have been used, including 1st-order derivatives (Roberts, 1965) and zero-sum filters, such as Prewitt's and Sobel's operators. The first order derivatives can best be described by assuming the image radiance values, r , represent the height value of a three-dimensional surface on an x,y coordinate plane:

$$r = f(x, y)$$

When y has a fixed value, the derivative will be determined by the difference between the value of x and $x + \Delta x$. The same procedure can be applied to a fixed value of x . The derivatives of the function for fixed x or y , the so called partial derivatives of r about x and y , are defined respectively as the following (Chiang, 1974):

$$f_x \equiv \frac{\partial r}{\partial x} \equiv \lim_{\Delta x \rightarrow 0} \frac{\Delta r}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

$$f_y \equiv \frac{\partial r}{\partial y} \equiv \lim_{\Delta y \rightarrow 0} \frac{\Delta r}{\Delta y} = \lim_{\Delta y \rightarrow 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

The first order derivative operators are based on this partial derivative concept, in which the gradient in the x direction is the derivative for the fixed value of y , and the gradient in the y direction is obtained for fixed x . Within the operators, the kernel values are divided into two groups on either side of the center row or column, and the difference between the two parts of the kernel is used to estimate the gradient. If the directions of the edges are unknown, or not regarded as important, then the average value of the partial derivatives about the x and y directions can be used as a proxy for the derivative at a point:

$$f = (f_x^2 + f_y^2)^{1/2}, \text{ or}$$

$$f = |f_x| + |f_y|$$

where:

f : The gradient

f_x and f_y : The gradient in x and y directions

The above approach, commonly used in traditional edge enhancement algorithms, tends to blur the representation edges through a smoothing effect. For the Roberts edge detection operator (Roberts, 1965) the gradients are determined by:

$$f_x = DN_{i,j} - DN_{i+1,j+1}$$

$$f_y = DN_{i+1,j} - DN_{i,j+1}$$

and the Sobel edge detection operator (Holyer and Peckinpaugh, 1989) is calculated from:

$$f_x = - DN_{i-1,j-1} - 2DN_{i,j-1} - DN_{i+1,j-1} + DN_{i-1,j+1} + 2DN_{i,j+1} + DN_{i+1,j+1}$$

$$f_y = - DN_{i-1,j-1} - 2DN_{i,j} - DN_{i,j+1} + DN_{i+1,j-1} + 2DN_{i+1,j} + DN_{i+1,j+1}$$

where:

i and j : The row and column values that define a pixel location

DN: Spectral digital values

In contrast to the techniques discussed so far, Laplacian techniques use an approximation of 2nd-order derivatives, which differentiate the first order derivatives for every pixel. The 2nd order derivatives have a zero value where the gradients reach a local maximum. The formula used is as follows:

$$f = f_x + f_0 + f_y$$

For the unweighted Laplacian operator (ERDAS, 1999),

$$f_x = - DN_{i-1,j-1} - DN_{i,j-1} - DN_{i+1,j-1}$$

$$f_y = - DN_{i-1,j+1} - DN_{i,j+1} - DN_{i+1,j+1}$$

$$f_0 = 2DN_{i-1,j} + 2DN_{i,j} + DN_{i,j} + 2DN_{i+1,j}$$

The intensity gradient methods are generally simple to apply, but tend to be too sensitive to noise, fine-structure, and weak gradients (Holyer and Peckinpaugh, 1989).

To limit these problems, normalization can be used. For example, Fjørtoft *et al.* (1998) used an infinite symmetric exponential filter (ISEF) as an edge detection operator to eliminate the effect of noise in a synthetic aperture radar image. The filter normalizes the pixel values with an ISEF that calculates the exponentially weighted means of each half window. The center pixel's gradient is computed based on the normalized values of the two half windows.

3-3. Image segmentation methods

Thresholding of DN values and split and merge methods are not well matched conceptually with object identification. Although region growing methods are more appropriate for object identification, they require user-selected levels for the similarity criteria. Furthermore, the results of region growing may be dependant on the starting point and order of region growing. Therefore, in this study an edge-based approach was used for image segmentation. However, it should be borne in mind that drawbacks of the edge-based approach include the requirement of a user-specified threshold for an edge to be regarded as significant, and the need for several additional procedures to produce segments from the edges. In this research the edge detection method consists of five steps: edge enhancement, ridge finding, ridge connection, ridge thinning, and ridge pruning (Figure 3-1).

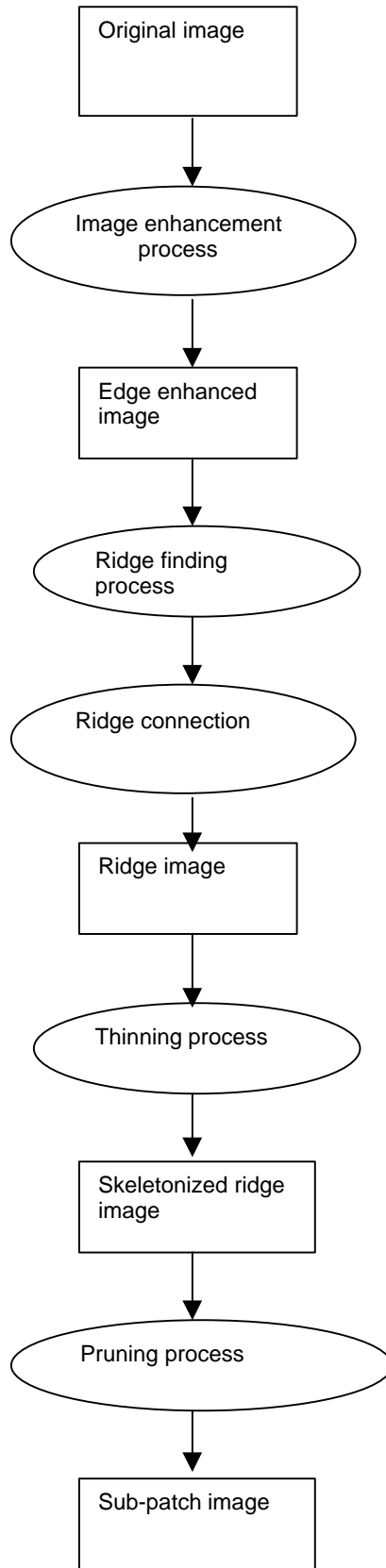


Figure 3-1. The image segmentation process of identifying sub-patches in the image.

3-3-1. Image enhancement

The image segmentation method used is an edge-based approach, which has been modified to incorporate directional linear kernels. The gradients around a point are assumed to have a maximum value in a direction perpendicular to an edge, if one is present. In order to exploit this approach, the partial first derivative concept should be substituted by the total first derivative concept, in which partial derivatives are extended to all directions with the simultaneous change of x and y . If r is a function of x and y , the total differential is thus the change of r by the change of x ($f_x dx$) plus the change of r by the change of y ($f_y dy$):

$$dr = f_x dx + f_y dy \dots\dots\dots(3-1)$$

When $y = h(x)$, the change of x affects not only r directly, but also y . In order to count the two effects, the total derivative is required. The total derivative of r with respect to x is expressed as follows (Chiang, 1974):

$$\begin{aligned} \frac{dr}{dx} &= f_x \frac{dx}{dx} + f_y \frac{dy}{dx} \\ &= \frac{\partial r}{\partial x} + \frac{\partial r}{\partial y} \frac{dy}{dx} \end{aligned}$$

When $y = h(x)$, the maximum total derivative g at (x,y) among the possible directions can be defined by:

$$g = \max \frac{dr}{dx} \dots\dots\dots(3-2)$$

In order to produce a proxy based on this approach, it is necessary to adapt the shape of the operators. Rectangular shaped operators used by traditional edge enhancement

operators are not appropriate for estimating the total derivative. Rectangular operators are insensitive to edge shape, specifically edges with sharp curves. As discussed earlier, the operators divide the kernels into two parts that are generally a half of the kernel, and produce gradients by comparing the difference between them. In the case of corner of an object, neither half of the kernel represents the object under investigation. This will tend to cause an offset in the location of the derived edges. This problem is likely to be most acute for the larger windows that may be particularly important with high spatial resolution data (Ferro, 1998).

In this study, directional linear operators are used for edge enhancement. Because linear shaped operators are only a single pixel wide, the operators will be affected less by non-straight edges. However, as the formula (3-2) implies, the operators are sensitive to directional differences. This is crucial in order to measure the magnitude of the gradient correctly. Figure 3-2 (a) shows the bright roof of a building from the ADAR data of Morgantown. Profiles in the green DN values for three directions at a point on the roof edge are marked A, B, and C in the image, and are graphed in Figure 3-2 (b). The profile that is perpendicular to the edge has the steepest slope, while the profile that is parallel to the edge has the gentlest slope.

The green band was selected for the edge enhancement and ridge finding process because this band was found to produce the best results. The red and blue bands, were found to have a lower contrast in the ADAR data, and therefore produced edge images with more spurious edges than the green band. The infrared band provided good discrimination of green vegetation, but did not discriminate the edges of important cover types.

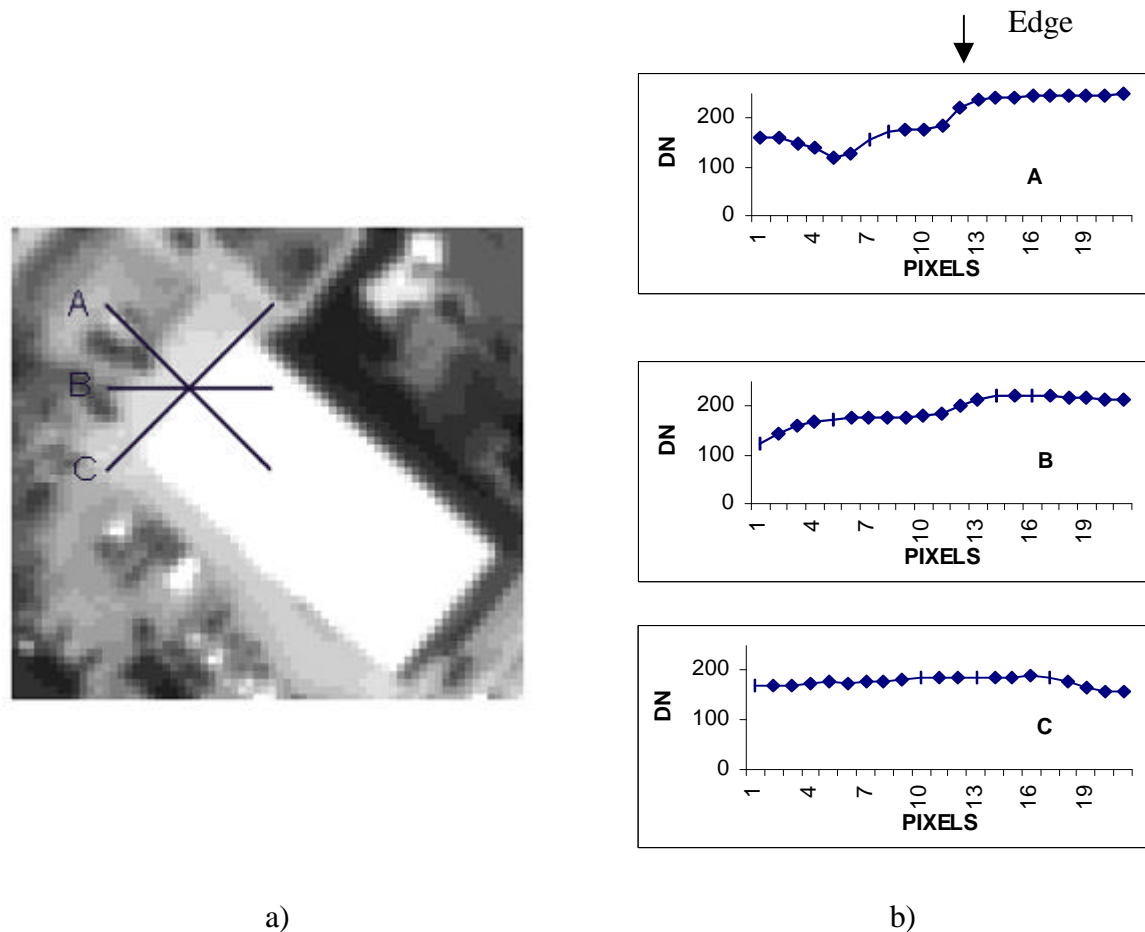


Figure 3-2. Spatial profiles for three directions for a subset of the green band of the Morgantown ADAR image. a) is the subset of Morgantown ADAR image, and b) is the spectral radiance profiles for the three directions.

The size of operator used is partly determined by the scale of the imagery. When situated on an object edge, each half of the kernel should be large enough to capture the spectral properties of the objects or background, but not too large to straddle more than one object.

The size of the operator also determines the number of potential directions that can be examined, with larger operators allowing more directions. For example, if a linear directional operator with five pixels is used, eight directions can be investigated (Figure 3-3).

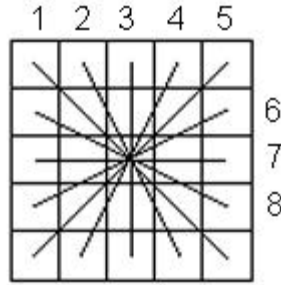


Figure 3-3. The eight directions associated with a linear operator five pixels in extent.

The use of the directional operator to identify the direction that is perpendicular to an edge is shown in Figure 3-4, which shows the gradient index calculated from the first

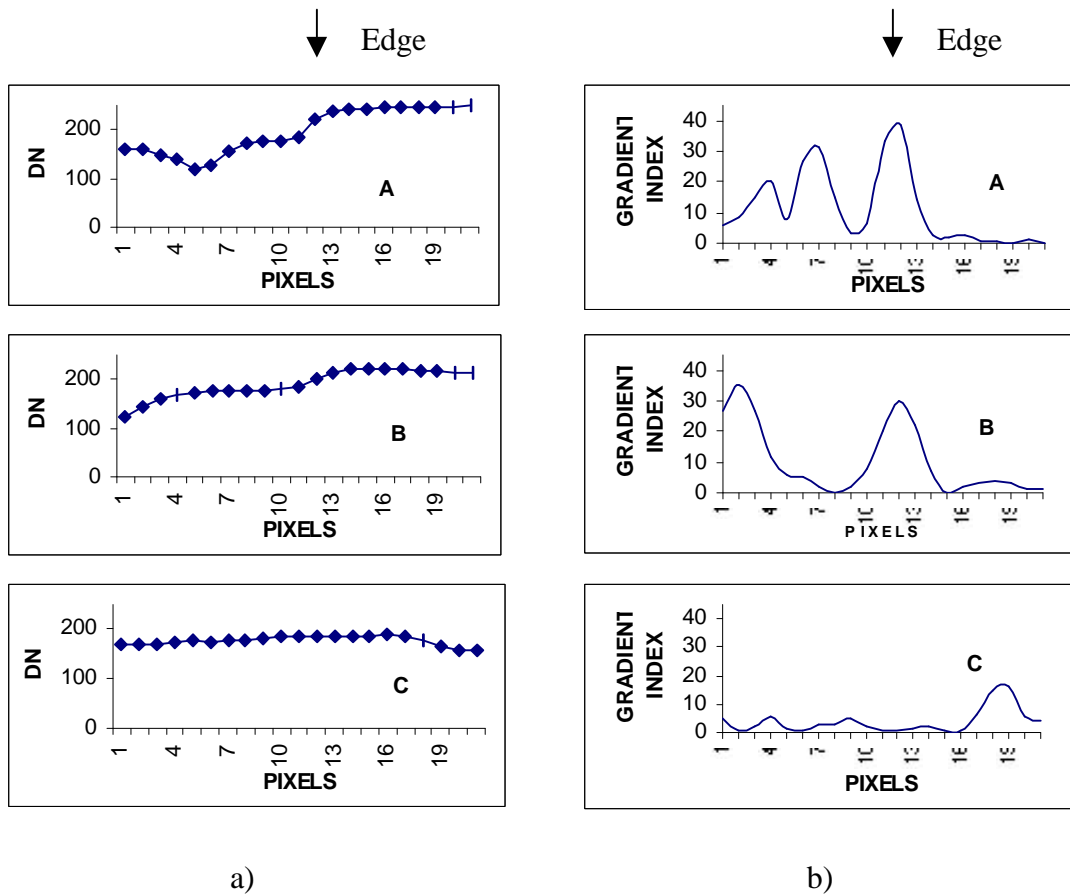


Figure 3-4. First order derivatives for three directions for the image shown in Figure 3-2. a) is the spectral radiance profiles for the three directions, and b) is the corresponding gradient index, which is calculated from the first derivative.

derivative of the three profiles illustrated in Figure 3-2. Profile A, which is perpendicular to the edge of the building, has the highest gradient measurement. By comparison, the direction parallel the building edge has the lowest values.

The sensitivity of the kernel shapes and sizes described in the previous section has been examined through the application of a variety of filters to simulated data sets (Figure 3-5) using custom-written Fortran programs. The data sets are composed of geometric objects that are distinguished from their background by a strong contrast in radiance values. The objects have three edge shapes, namely perpendicular (A), round (B) and acute (C). These shapes were chosen because perpendicular edges are typical of buildings, whereas all three types of edges are associated with roads. Three operator sizes were tested: 3, 7, and 11 pixels. In Figure 3-5, the upper case letters indicate the results of gradient measurements from directional linear operators, and the lower case letters indicate the results from traditional square shaped operators. Firstly, the linear operators provide a narrower, and more consistent delineation of edge location than the square shaped operators (compare the rows labeled A, B and C to those labeled a, b, and c). The advantage of the linear operators is most clear in the delineation of the acute angled corners of the triangles (C-3 vs. c-3). Secondly, the blurring effect of larger operators (compare columns 2 and 3 to column 1) is much less significant for the linear operators, compared to square operators. With square operators, the highest texture is found when the operator is displaced from the corner, because when the operator is shifted towards the interior of the object, more edge pixels are within the operator.

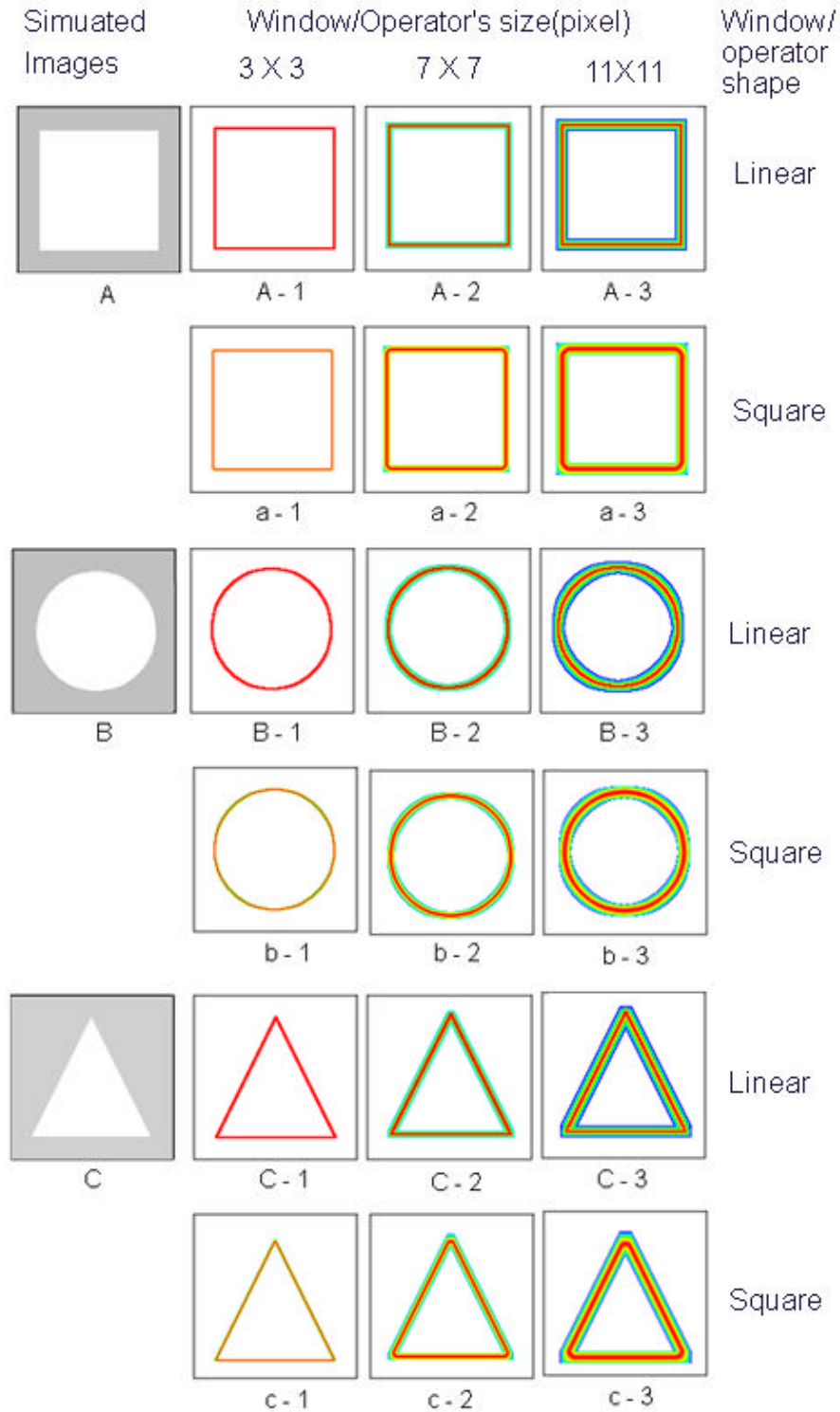


Figure 3-5. A comparison of edge enhancement images for three simulated images (A, B, and C) using a directional linear operators and traditional rectangular shaped kernels. The colors in the edge detection image are coded according to the strength of the response: from blue (low) to red (high).



Figure 3-6. Edge enhancement of Morgantown ADAR data for a subset of the study area, including downtown and parts of the neighboring suburbs.

Because the minimum size of an object of interest in this study is approximately two to three meters, a three pixel operator was used for the edge enhancement in this study. If a size of operator greater than this is used, spectral mixing between multiple classes or objects is likely. Figure 3-6 shows the results of the edge detection applied to the ADAR data of Morgantown. The edges of buildings are clearly delineated in light tones (high values), however, the edges of roads and vegetated regions are less well defined.

3-3-2. Ridge finding

The result of edge enhancement is a continuous tone image. Edge extraction is then applied to this image in order to select those pixels that best represent edges. For this study, it was assumed that edge pixels are represented in the edge enhancement image by the local highs, termed ‘texture ridges,’ or ‘ridges’ for short. Ridges can be identified by values that are higher than adjacent pixels in at least one transect through the point of interest. Thus no *a priori* window size for maximum ridge width need be specified. Both peaks and saddles in the edge-enhanced data are classified as ridges. An important aspect of this procedure is that the transects are extended until a local maximum is reached, and therefore even wide flat areas can potentially be identified as ridges. Transects are searched in four directions: rows, columns and both diagonals.

This ridge finding procedure can be expressed more formally as the identification of a pixel or pixels that locally have a zero value total derivative, but that on one side are immediately adjacent to a region with a positive total differential, and on the other side

are adjacent to a region with a negative total differential. Furthermore, the positive and negative total differentials should be co-linear in direction.

The points that have these properties are the edges in the spectral reflectance measurement space. This means that the derivative of equation 3-1 for the direction should have zero value. This result can be shown to be a consequence of differentiating equation 3-1 once again:

$$\begin{aligned}
 d^2 r &\equiv d(dr) = \frac{\partial(dr)}{\partial x} dx + \frac{\partial(dr)}{\partial y} dy \\
 &= \frac{\partial}{\partial x}(f_x dx + f_y dy) dx + \frac{\partial}{\partial y}(f_x dx + f_y dy) dy \\
 &= (f_{xx} dx + f_{xy} dy) dx + (f_{yx} dx + f_{yy} dy) dy \\
 &= f_{xx} dx^2 + f_{xy} dy dx + f_{yx} dx dy + f_{yy} dy^2 \\
 &= f_{xx} dx^2 + 2f_{xy} dx dy + f_{yy} dy^2 = 0
 \end{aligned}$$

A Fortran edge extraction program was written to implement the procedure outlined above. Because of the simplification required for raster data, junctions of ridges were sometimes found to have discontinuities. This problem was overcome with a ridge connection procedure that checked for one-pixel gaps between ridges with differing directional trends.

Figure 3-7 shows the results of applying the ridge finding process to the subset first shown in Figure 3-2. In Figure 3-7, it can be seen that spectral contrast in the scene gives rise to texture ridges along the building edges, as well as around many other fine scale objects in the scene. However, many of the ridge pixels are in clumps several pixels wide in areas where the DN values are similar along an edge. This occurs because the

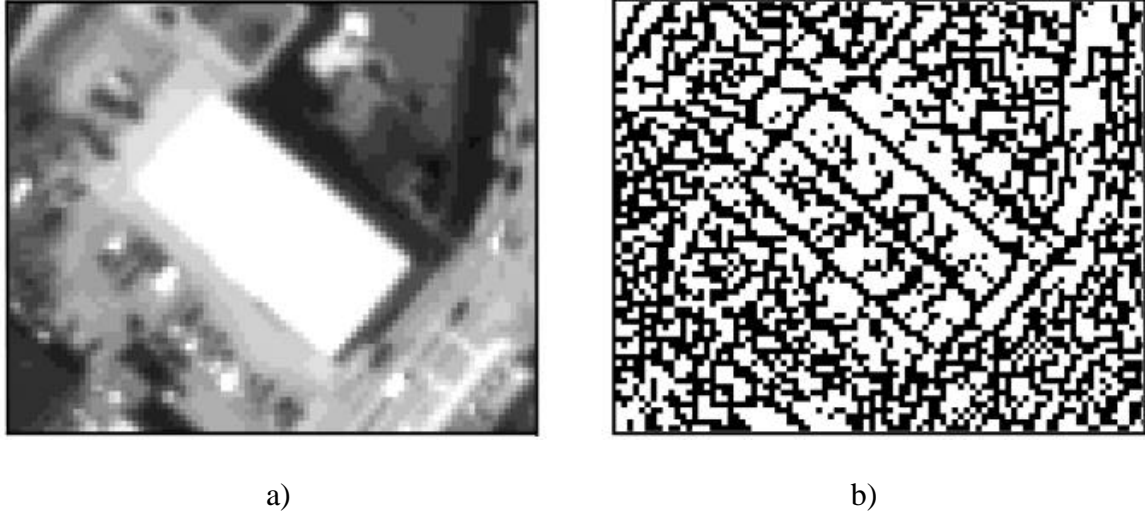


Figure 3-7. A small test subset of the ADAR data of Morgantown, and the results of the ridge finding process. a) is the original green band image, and b) is the results of ridge finding

transects are oriented in multiple directions, and thus spurs on the ridge are also defined as ridges.

3-3-3. Thinning

The ridge pixels were skeletonized into lines one pixel wide. The thinning procedure was implemented iteratively by erosion from left to right, right to left, downward, and then upward. In the thinning process, each pixel was checked to see if removing it would break the continuity of a polygon edge. Only pixels that are redundant for defining polygons were removed. Where an edge that had two or more pixels wide was found, the ridge pixels with the lowest edge enhancement value was eroded. The procedure was designed to retain ridge pixels over the original edges in the spectral

reflectance measurement space. Figure 3-8 (a) shows the results after applying this process.



Figure 3-8. The results of the thinning and pruning process. a) Skeletonized image after thinning process, and b) Image after pruning process for a part of the study area. (Compare with Figure 3-7.)

3-3-4. Pruning

The thinning process results in skeletonized ridges that have many branches that terminate without forming closed polygons (Figure 3-8a). Branches tend to represent edges that do not define object boundaries. Based on a careful examination of the image, removing these branches does not appear to remove any real edges, and therefore branches can be removed. This was achieved through an iterative erosion of pixels that are terminal nodes on the skeletonized ridges. This leaves only ridges that define closed polygons (Figure 3-8b), termed sub-patches in this work. The sub-patches represent segments of relatively homogeneous radiance within an object, and are bounded by zones

of the spectral variation. This approach is necessary, because as Mason *et al.* (1988) has noted, segmentation errors or ambiguities are inevitable. Errors can arise from noise in the scene, or real variations within the individual objects, few of which are completely homogeneous. The river and vegetated areas tend to be highly segmented through the pruning process, presumably due to variation in the tree and wave distributions of these two classes.

3-4. Application to simulated and real data

Figure 3-9 shows the result of the ridge finding procedure applied to the simulated data (Figure 3-5). The first image in every row represents the overlapped images of ridges identified by the different window sizes in the row. It is clear that the linear shape edge detection operators produced stable and accurate estimates of the edges of the geometric objects. Larger window sizes did not diminish the accuracy of the estimate of the edge location, except for a slight offset of the triangle corner using the seven pixel linear window. In contrast, the traditional rectangular edge detection operators produced results that were less accurate than the linear operator, even for the smallest window (3 by 3). Furthermore, results for the rectangular edge detection operators deteriorated with larger window sizes.

Figure 3-10 shows the image sub-patch segmentation results, using a directional linear operator, for a subset of the Morgantown ADAR data. The river and vegetated areas tend to be highly segmented through this process, presumably due to variation in

the tree and wave distributions of these two classes. In the next two chapters, the procedure to merge sub-patches to form patches, and ultimately objects, is described.

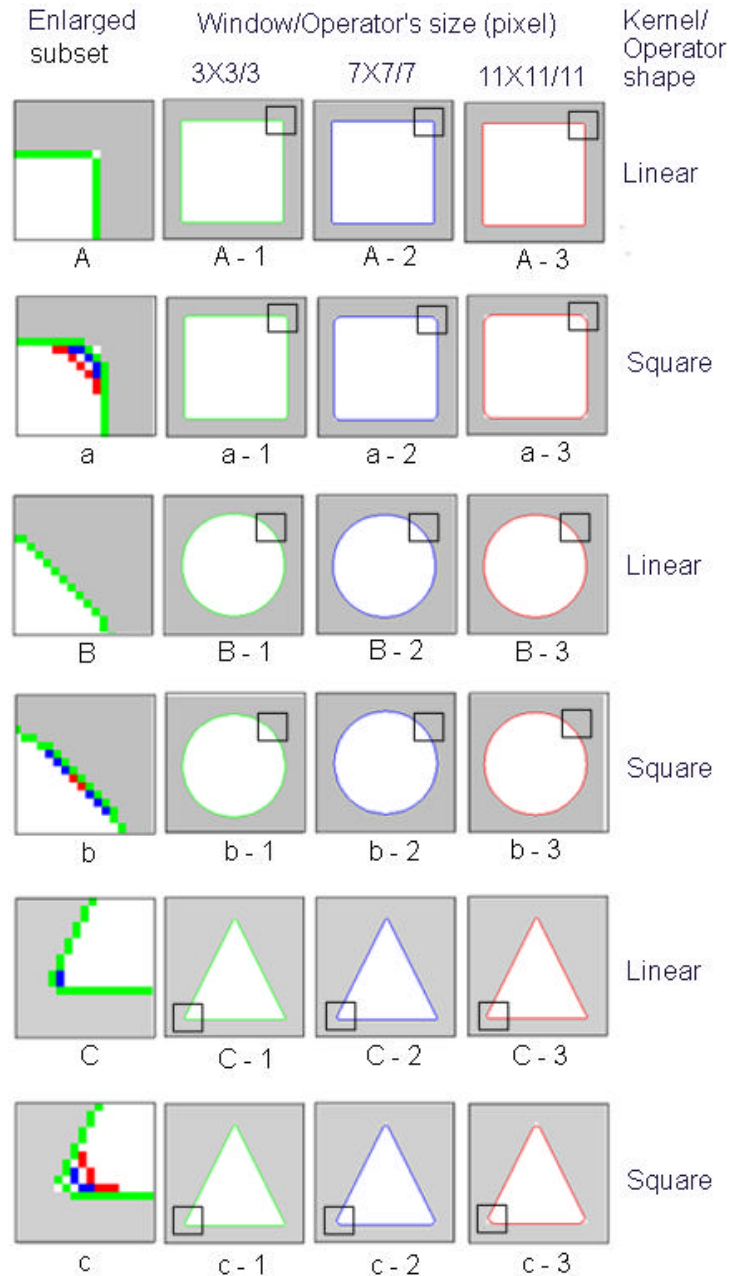


Figure 3-9. Comparison of edge detection results for directional linear (rows with upper case letters) and square operators for three sizes of windows/operators: 3 by 3/3 (green), 7 by 7/7 (blue) and 11 by 11/11 (red). The leftmost column shows an enlarged composite of the three edge detections, with the smallest windows overlain on the larger. Compare to Figure 3-5.



Figure 3-10. The sub-patch boundaries identified in the initial image segmentation stage of the Morgantown ADAR data overlain on a standard false color composite of the scene.

Chapter IV

REGION GROWING WITH IMAGE SEGMENTS

4-1. Introduction

Edge detection methods tend to be sensitive to noise and slight variations in reflectance even within objects (Holyer and Peckinpaugh, 1989). Thus when edge detection is used for image segmentation, the delineated areas are often much smaller than a single object. It is referred to these delineated regions as sub-patches in order to emphasize the necessity to amalgamate the initial segmentation into spatial units that can be related to image objects. The presence of these overly segmented sub-patches within the image objects reduces the effectiveness of object identification. A major issue in edge detection is therefore the discrimination of true edges from spurious edges (Fjørtoft *et al.*, 1998). Unfortunately there is no general method to achieve this (Elder and Zucker, 1998).

One technique used to distinguish significant edges, is to threshold only the higher values after edge enhancement (Lea and Lybanon, 1993; Karniell *et al.*, 1996; Touzi *et al.*, 1988). For example, Fjørtoft *et al.* (1998) was able to find a satisfactory threshold

that greatly reduced the over-segmentation produced by their edge-finding, local maxima watershed algorithm (Vincent and Soille, 1991). Generally, however, selecting a threshold is not necessarily a strait forward process because the optimal threshold tends to vary across the scene. Adjacent low contrast objects require a very low threshold, whereas in regions of high contrast a much higher threshold is required (Elder and Zucker, 1998; Fjørtoft *et al.*, 1998). If the threshold is set too low, many spurious edges are identified. However, if the threshold is set too high, additional complex processing is required to subdivide segments.

The four criteria of Haralick and Shapiro (1985) for ideal image segmentation help explain why the selection of edges that define segments is so difficult. First, segmentation should result in the grouping of pixels with uniform or homogeneous characteristics, such as gray tone or texture. Second, adjacent regions of a segmented image should have different radiometric or texture characteristics. Third, each region should have spatially accurate and simple boundaries. Fourth, the interiors of the regions should not have numerous small holes. Finally, it is also important to satisfy all the criteria simultaneously (Haralick and Shapiro, 1985).

In this research, the first criterion of internal homogeneity was addressed by applying thresholding and region growing methods sequentially (Figure 4-1). An optimal threshold level was selected by incorporating a dual approach in which vegetation boundaries were selectively identified in a separate procedure. For the second criterion, that adjacent regions should be different, the spectral similarity of adjacent sub-patches was evaluated by multivariate statistical methods, and sub-patches that did not have sufficiently different radiometric values were merged. The third criterion, relating to

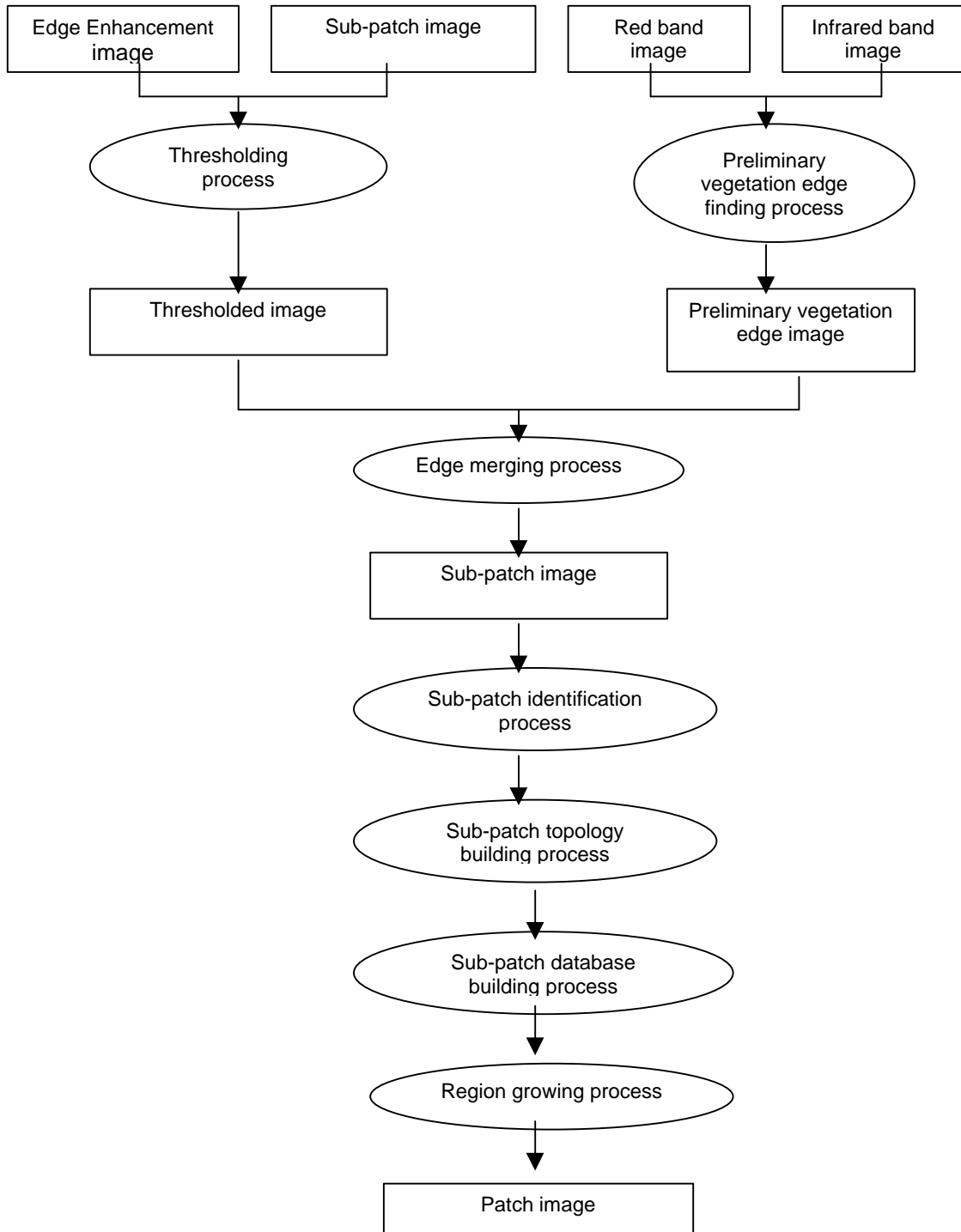


Figure 4-1. Overview of the region growing stage and patch identification.

simple and accurate boundaries, was satisfied partly by the ridge finding technique described in Chapter III, and is refined further by the region growing described in this chapter. The fourth criterion, specifying that holes within regions are undesirable, was addressed in two stages. Firstly, many of the small holes within objects were removed by the thresholding procedure itself. Secondly, small unwanted objects, such as cars on roads, and small structures on the roofs of buildings, were retained until after the classification stage, and only then merged with the surrounding classes. The reason for keeping such objects until after classification is that they tend to have distinctive spectral characteristics, and thus their inclusion in the surrounding class would violate the first two criteria of Haralick and Shapiro (1985).

This chapter describes the procedures for identifying edges, the thresholding procedure, the evaluation of the similarity of adjacent sub-patches, and the aggregation to form patches (Figure 4-1). In Chapter V, the procedure for patch classification and the delineation of objects, is described.

4-2. Thresholding

An empirical approach was used to select the optimal threshold for identification of the sub-patch edges from the edge enhancement data. The binary ridge data was converted into continuous data by multiplying the ridge finding result and edge enhancement data. A threshold was applied to the resulting image, in order to suppress low edge enhancement value ridges. After applying the threshold, a cleaning process is carried out to remove the remaining sections of polygon edges that have been broken in

the thresholding process. Real edges can potentially be lost through this cleaning, because no test is applied to determine what proportion of the edge has been removed by applying the threshold. Nevertheless, this procedure appears to remove very few real edges because a very low threshold is chosen.

A number of thresholds were tested, and the derived edges were overlaid on the original image for visual comparison. For example, Figure 4-2 demonstrates the differences in segmentation that result from applying varying thresholds to the output of the ridge finding process. When the threshold level is set to 30 (units are arbitrary edge enhancement DN values), the building in the center of the image is segmented into many small sub-patches; however, with a threshold level of 150, the objects in the image are

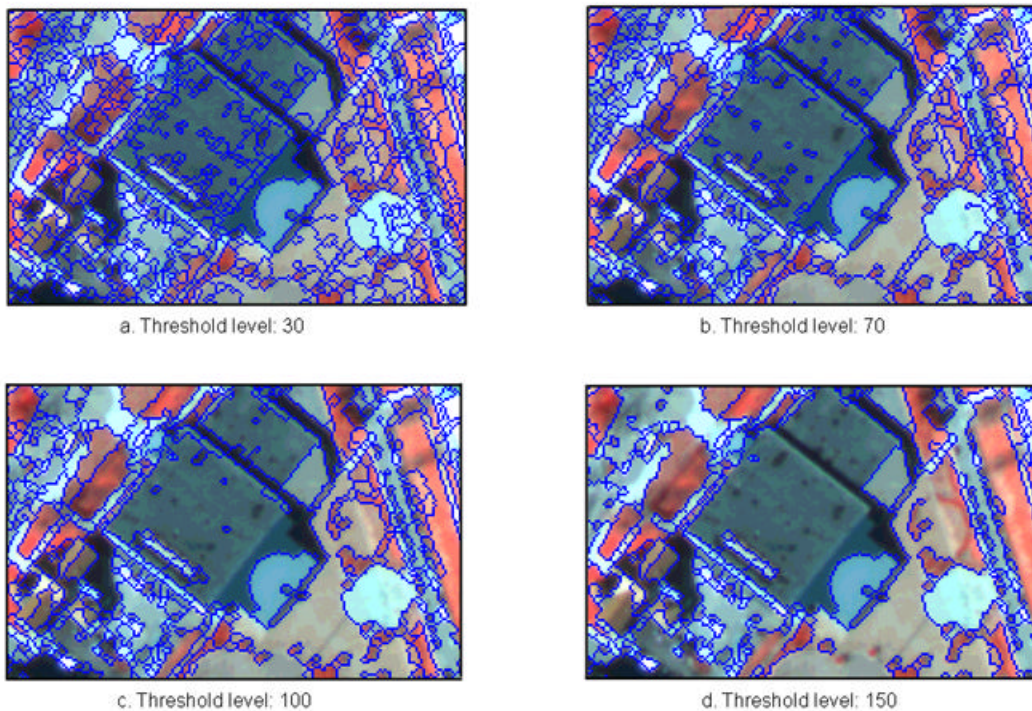


Figure 4-2. Comparison of threshold levels of 30, 70, 100, 150 for image sub-patch boundary identification for a small test site within the Morgantown data.

grouped into large sub-patches that have an undesirably wide range of radiance values.

The final threshold level was therefore chosen to minimize the grouping of separate objects into a single clump, with the exception that vegetation boundaries were ignored because they could be identified separately, as is described in Section 4-3. Thus, a level of 70 was selected for the final threshold, despite the aggregation of vegetation and other cover classes within individual sub-patches (Figure 4-3). Although the separate treatment of vegetation boundaries allowed a much higher threshold than would otherwise have been possible, a threshold of 70 results in many small sub-patches. Some of the small sub-patches are a result of cars on the roads and structures on the roofs of buildings. The forest area also has small sub-patches produced by the geometric patterns of trees and tree stands. However, for the water class, the threshold is sufficiently high that the entire Monongahala River was segmented into a single sub-patch because ridges found in the water class (Figure 3-10) were all below the final threshold chosen.

4-3. Preliminary extraction of vegetation edges

Vegetation areas were found to have relatively diffuse boundaries and a low contrast with adjacent land cover types such as road, water, and shadow. This was a consequence of the selection of an early spring image for the analysis. However, a summer image would have had additional problems associated with the obscuring effect of vegetation on ground objects. Therefore, vegetation edges required a very low threshold in this image. Nevertheless, because vegetation has such a distinctive spectral signature, it is possible to identify vegetation edges through an alternative approach, namely a rule-based pre-



Figure 4-3. Image sub-patches delineated by an edge enhancement threshold level of 70.

classification. A number of rule-based classifications have been found to be effective. For example, Ton *et al.* (1991) developed rules that use a standard TM vegetation index, the ratio of infrared to red (Band 4/Band 3): Pixels with a vegetation index greater than 2.5 were labeled as vegetation, below 1.5 as non-vegetation, and between 2.5 and 1.5 as ambiguous. In a slightly more complex approach, Møller-Jensen (1990) used an approach based on the expert system developed by Wharton (1987) for the identification of green vegetation with Landsat TM data. In Møller-Jensen's method, vegetation pixels must have (band 4 + band 5) > (band 2 + band 3 + band 7), and band 4 > band 5.

Because the bands used in this study were different from those of Landsat TM, and because this image was acquired prior to forest leaf out, new rules exploiting the spectral characteristics of green lawns and senesced deciduous forest vegetation were developed. In the first step, a normalized difference vegetation index (NDVI) (Tucker *et al.*, 1981), was employed for identification of potential vegetation:

$$NDVI = \frac{Band4 - Band3}{Band3 + Band4}$$

where:

band 3: ADAR red band

band 4: ADAR infrared band

The initial rule was that pixels with NDVI > 0.0 were identified as Potential Vegetation. For the second step, Potential Vegetation pixels that fell in the range from 100 to 140 in band 3, and 115 to 150 in band 4, were labeled Potential Non-Vegetation. In the third step, Potential Non Vegetation pixels were reassigned back to the vegetation if band 1 < band 2 < band 3 < band 4 (i.e. the DN values increase across the bands

sequentially). Figure 4-4 presents the results of the application of this procedure to the Morgantown ADAR data.

The vegetation edges were obtained directly from the boundary pixels of the vegetation areas, rather than applying the edge enhancement method used for other edges. Notice in Figure 4-4 that the lower-left part of image has a very simple structure, because this area is highly built-up, and does not have many green spaces. However, a number of small green areas were detected. The upper-right side of the image is a residential zone with a forested area along a stream, and consequently has rather complex vegetation edges.

The edges from the results of the previous two procedures were combined into one edge-detection image. After pruning edges that terminate without forming a closed polygon, the sub-patch image is complete (Figure 4-5). A careful comparison with Figure 4-3 shows that roads, water, and some shadows, are delineated better in Figure 4-5 than in the initial segmentation of Figure 4-3, which only incorporates the general edges from the ridge-finding procedure.

4-4. Sub-patch topology and database development

Image segmentation results in the labeling of each pixel according to the spatial grouping into which it falls. This data can be used for additional image processing analysis (Woodcock and Harward, 1992), such as amalgamation to form entities that can be related to image objects. In this work, sub-patches are treated conceptually as proto-objects, with spatial topology that goes beyond the traditional pixel-based segmentation.



Figure 4-4. Preliminary vegetation areas and associated edge pixels. Edge pixels are depicted in black, vegetation in green. Compare to Figures 4-3 and 4-5.



Figure 4-5. Image segments of sub-patches modified by vegetation edges.

The research of Gahegan and Flack (1999), Usery (1996), Albrecht (1996), and Roberts and Gahegan (1991) provides a conceptual framework for the analysis of objects or proto-objects in remote sensing. Roberts and Gahegan (1991) considered features to have associated spatial and aspatial properties. In Usery's (1996) approach, objects are geographic entities that incorporate spatial, thematic, and temporal information. Albrecht (1996) defined real-world geographic entities as geographic objects that have three components: geometric, metadata, and a set of relationships with other objects. A common theme with these approaches is that the object comprises more than the geometric properties, but includes geographic representation. (Roberts and Gahegan,1993).

This research builds on this geographic approach to object definition. A Fortran program was developed to provide the database for cataloguing and accessing the spatial and aspatial properties of the sub-patches, as well as for the subsequent processing. The analysis is therefore based on sub-patches as the spatial unit, rather than pixels, and the spatial relationship between the sub-patches are an important part of the processing. The irregularly shaped sub-patches were treated as raster polygons, which are similar in concept to GIS polygons. These raster polygons, however, were defined by pixels instead of the points and lines that are used in a vector-based GIS. Nevertheless, the relationships between adjacent polygons can be defined by the raster network, as the sub-patch edge pixels can be regarded as the arcs of a traditional vector GIS. Figure 4-6 shows an example of the sub-patch identification for a part of the study area. Sub-patches are the white areas separated by black sub-patch edge pixels. The sub-patch identification numbers were assigned by a rule developed in this study.

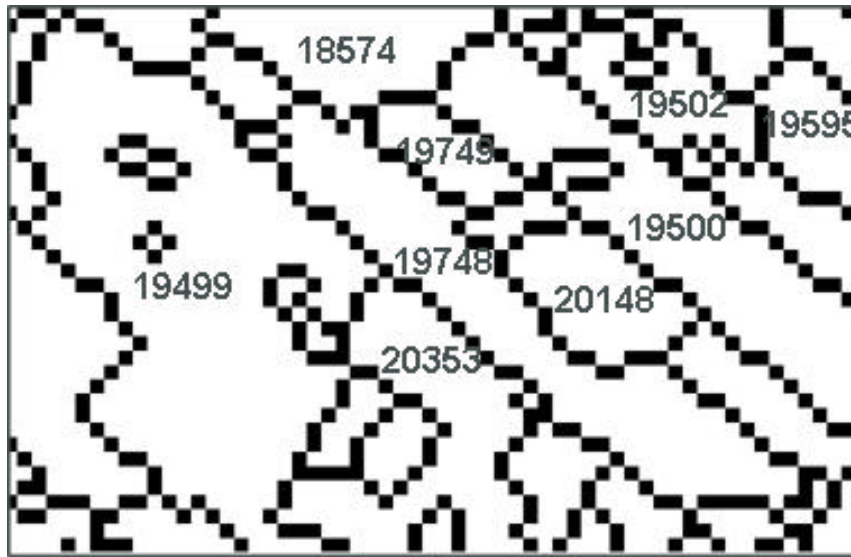


Figure 4-6. An example of the sub-patch segmentation, and the associated polygon identification codes for selected polygons.

A database was built for the sub-patches, including the topology and statistical parameters based on the distribution of spectral radiances of the pixels within each sub-patch. The database includes four main information sections for each sub-patch: the polygon identification information, spectral measurements, topology, and aspatial information. Identification information includes each sub-patch polygon identification number, as well as the locations of sub-patch edge pixels and the pixels that belong to each sub-patch. Spectral information includes the original DN values of the pixels in the sub-patch. Topology contains the spatial relationships between contiguous sub-patches. Aspatial information incorporates the number of pixels in each sub-patch and the average DN value, sum of squares, variance for each band, as well as the covariance between each possible pair of bands. This information provides the basis for merging the sub-patches, as well as later classification.

4-5. Region growing with image segments

In the process of segmentation, it was assumed that it is easier to merge sub-patches than to subdivide them. Therefore, the threshold chosen resulted in many sub-patches within objects. In the next processing step, the method to identify and merge similar sub-patches is explored.

The region growing procedure is based on a test for similarity between adjacent sub-patches. This procedure differs in two aspects from generic region growing methods that merge regions sequentially. Firstly, the similarity is quantified for all pairs of spatially adjacent sub-patches identified by the topological relationships in the database, and the merging procedure is subsequently applied in a single pass through the data. This approach, in which all adjacent regions are compared prior to any merging of regions, has the advantage that the starting point of region growing does not determine the result. Furthermore, the fact that the sub-patches are not arbitrary units, such as a square window, but instead are regions with similar spectral characteristics, allows a more flexible definition of the seed units for amalgamation..

The second important difference compared to previous studies is the way in which multivariate statistical analysis was used in the region growing. Several criteria for merging segments have been used, including minimum distance, likelihood ratio (Kettig and Landgrebe, 1976; Oliver *et al.*, 1996), and student's *t* test (Fjørtoft *et al.*, 1998). Kettig and Landgrebe (1976) also used a "multiple univariate approach" that did not consider the covariance between variables for sub-patches.

The approach used in this study was a multivariate form of univariate t^2 statistics. With univariate t^2 analysis, the means from two groups are assumed to be drawn from a single population. The null hypothesis is that the two means are similar, and a certain level of significance is associated with the result. If the null hypothesis is proved, it is accepted that the two groups originated from the same population.

Let $X_i=(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in_i})$ represent the elements of a group X_i . Then the similarity between two groups X_1 and X_2 can be tested with t^2 :

$$t^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}} \dots\dots\dots(4-1)$$

Where:

- s_1^2 : variance of group X_1
- s_2^2 : variance of group X_2
- \bar{X}_1 : mean of group X_1
- \bar{X}_2 : mean of group X_2
- n_1 : number of observations of group X_1
- n_2 : number of observations of group X_2

Equation (4-1) can be extended to multivariate data (Anderson, 1984). Suppose that there are two groups of observations with multiple variables. The two groups X_1, X_2 have respectively n_1, n_2 number of observations, for p variables. Then the observation of group X_i , for $i=1,2$, can be represented in matrix form:

$$X_i = [x_{ijk}] = \begin{bmatrix} x_{i11} & x_{i12} & \dots & x_{i1p-1} & x_{i1p} \\ x_{i21} & x_{i22} & \dots & x_{i2p-1} & x_{i2p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{in_i-11} & x_{in_i-12} & \dots & x_{in_i-1p-1} & x_{in_i-1p} \\ x_{in_i1} & x_{in_i2} & \dots & x_{in_ip-1} & x_{in_ip} \end{bmatrix}$$

where:

i : i th group

j : j th observation in group i

k : k th variable

n_i : number of observation in group i

p : number of variables involved

x_{ijk} : value for j th observation for k th variable in i th group

The scalar mean for univariate analysis is replaced by a mean vector, \bar{X}_i , for multivariate analysis.

$$\bar{X}_i = \begin{bmatrix} - \\ x_{i1} \\ - \\ x_{i2} \\ \dots \\ - \\ x_{ip-1} \\ - \\ x_{ip} \end{bmatrix}$$

The scalar variance for the univariate case is extended to a variance-covariance matrix, S_i , for multivariate analysis:

$$S_i = \begin{bmatrix} S_{i11} & S_{i12} & S_{i13} & \dots & S_{i1p} \\ S_{i21} & S_{i22} & S_{i23} & \dots & S_{i2p} \\ S_{i31} & S_{i32} & S_{i33} & \dots & S_{i3p} \\ \dots & \dots & \dots & \dots & \dots \\ S_{ip1} & S_{ip2} & S_{ip3} & \dots & S_{ipp} \end{bmatrix}$$

Where:

s_{ij} : variance of jth variable for the group i

s_{ijk} : covariance between jth and kth variables for group i

The multivariate equivalent of equation (4-1) can now be defined as:

$$T^2 = \frac{(\bar{X}_1 - \bar{X}_2)^2}{S_1/n_2 + S_2/n_1} \dots\dots\dots(4-2)$$

This formula does not require that the two variance-covariance matrices be equal. However, the null hypothesis assumes that the variance-covariance matrices of the two populations are equal. If the null hypothesis is true, it is better to use all the data to estimate the common variance-covariance matrix, instead of first calculating the variance-covariance matrices for each of the two groups. When the two groups have a different number of cases, the common variance is best estimated using a weighting based on the number of observations of the two groups. This is intuitively correct for the diagonal elements of the variance-covariance matrices, which are simply the variances of each variable. Using the common variance-covariance matrix, the denominator of equation (4-2) becomes:

$$S = \frac{(n_1 - 1)S_1}{(n_1 + n_2 - 2)} + \frac{(n_2 - 1)S_2}{(n_1 + n_2 - 2)}$$

Where:

S : the pooled variance-covariance matrix

Using the common variance-covariance matrix, the denominator of equation (4-2) becomes:

$$S_1/n_2 + S_2/n_1 = \left(\frac{n_1 + n_2}{n_1 n_2} \right) S$$

Formula (4-2) then becomes:

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{X}_1 - \bar{X}_2)' S^{-1} (\bar{X}_1 - \bar{X}_2)$$

This T^2 -statistic, originally described by Hotelling (Anderson, 1984), has the following F distribution multiplied by the coefficient when \mathbf{m} is the population mean:

$$T^2 \sim \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} F_{(p, n_1 + n_2 - p - 1)} \dots\dots\dots(4-3)$$

When equation (4-3) is rearranged to focus on F , the result is as follows:

$$\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 \sim F_{(p, n_1 + n_2 - p - 1)}$$

If we find the T^2 value with multivariate data for two sample groups, the significance of the similarity of the groups can be examined by an F distribution with degrees of freedom p and $n_1 + n_2 - p - 1$.

4-6. Application to Morgantown ADAR data

In order to examine the performance of the proposed method, $\frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$ values were computed for all pairs of adjacent sub-patches identified in the topology database for the Morgantown ADAR data. The Hotelling T^2 test was based on all four bands of ADAR data, as well as NDVI. NDVI was included in the analysis for two reasons. Firstly, NDVI is very effective for vegetation discrimination, as discussed

earlier. Secondly, NDVI is a ratio measure, which tends to normalize for illumination variation, thus minimizing the effect of slope-aspect differences between sub-patches. The significance of the F distribution was calculated with the subroutines FTEST and AVEVAR, and functions BETACF and GAMMLN of Press *et al.* (1999).

Various levels of significance were tested for identifying similar sub-patches. Figure 4-7 shows the results for a small test site within the study area. The center of the image is a building with a complex structure. The image is oriented with north at the top of the page. The building has entrances on its northwest and southeast sides, and is surrounded by grass areas that are shadowed on the southwest and northeast sides, and part of the southeast side. A road is located on the southeast of the building, adjacent to the grass. The building roof has eight separate parts that have different elevations and which are separated by associated shadows. Each part was segmented into one or more sub-patches prior to the merging step.

Figure 4-7 shows that the lower the significance level used, the greater the degree to which the sub-patches are merged. The central part of the building had twelve sub-patches before region growing. When region growing with significance level of 0.001 (Figure 4-7e) was used, most of sub-patches of the central section of the building was merged into one patch. At the significant level of 0.00075, the road was amalgamated into a single patch.

Figure 4-8 shows the region growing results with significance level of 0.0005 for Hotelling multivariate T^2 test applied to the Morgantown ADAR data. Figure 4-8 shows a great improvement in segmentation over Figure 4-5, especially for roads, the forested area, and for some buildings. However, many small patches smaller than five pixels are

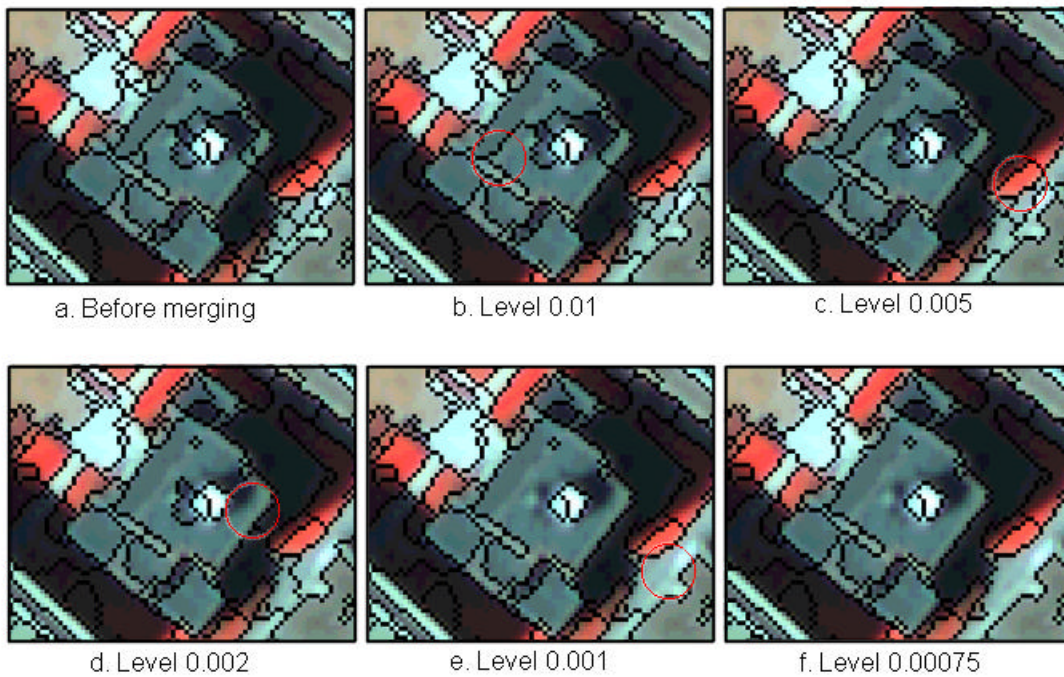


Figure 4-7. Comparison of different significant levels applied to the region growing procedure.

still present at this stage, as well as patches that appear to be zones of differing illumination caused by varying geometric structure within an object. These regions are dealt with in the next stage of the processing.



Figure 4-8. Patches produced by region growing with a significance level of 0.0005.

CHAPTER V

IMAGE CLASSIFICATION WITH A REGION- BASED APPROACH

5-1. Introduction

Perfect classification could be achieved if each spectral class were to have a unique spectral signature, because a one-to-one relationship could be established between spectral radiance pattern and class type. However, spectral overlap between most real classes occurs as a result of noise in the system, the natural variability of objects within a specific class, and the spatial variability of radiance within each object (Swain and Davis, 1978; Price, 1994). For example, because most tree species have similar leaf pigments, different tree species tend to have overlapping spectral properties. An added complication is that the spectral structure of an image is a function of scale (Cao and Lam, 1997). Higher spatial resolution may actually lead to greater variability within classes, as additional detail is resolved.

A number of studies have been carried out to find an appropriate spatial resolution to which to aggregate data in order to reduce the variation within an object, and minimize

the classification error (Pax-Lenney and Woodcock 1997; Teillet *et al.*, 1997; Latty *et al.*, 1985). Such approaches are pixel-based, and do not draw on the spatial variability as a source of information. Another problem with methods that search for an optimal scale is that real objects and classes are variable in size, and thus there is usually no single spatial resolution that suppresses all unwanted spectral variability (Marceau *et al.*, 1994a; 1994b). Studies that use image segmentation to identify single objects (Gougeon, 1995a) can overcome this problem of a single optimal scale. However, most such studies use mainly aggregated information such as average DN, and to a limited extent the variance within the image segments (Kettig and Landgrebe, 1976; Gougeon, 1995a; Meyer *et al.*, 1996). The variability within an object can provide additional information that can be used for image classification. The spectral correlation between bands, as quantified by the covariance matrix, is in fact often a key determinant in traditional maximum likelihood classification for separating classes that overlap in their univariate distributions.

In the next section, several traditional classification methods are reviewed. A new approach for classification of a previously segmented image is then described. The ADAR data, discussed in Chapters III and IV, is used to compare the new approach with traditional methods.

5-2. Traditional image classification

Pixel-based image classification procedures have focused on categorizing pixels in meaningful groups that have spectral homogeneity (Lillesand and Kiefer, 1994), or that

minimize the probability of misclassification (Swain and Davis, 1978). It is beyond the scope of this study to explore all of the many methods that have been developed. However, among the most commonly used methods are Iterative Self-Organizing Data Analysis Technique (ISODATA) (Tou and Gonzalez, 1974) for unsupervised classification and maximum likelihood for supervised classification. Another important image classification approach is the field-based method called ECHO (Kettig and Landgrebe, 1976). This latter procedure is available as part of the free software package, Multispec (Landgrebe and Biehl, 2000). Because this method employs some of the same statistical approaches one of the methods used in this study, ECHO will also be discussed in the next section.

The unsupervised ISODATA approach is a clustering algorithm (Jensen, 1996). The user specifies parameters for merging and splitting classes, as well as the number of initial clusters, which are normally spaced in some arbitrary fashion within the data space. Each pixel is initially assigned to the closest cluster based on a Euclidean measure of the spectral distance. Each cluster mean is then recomputed from the pixels that have been assigned to the cluster. Using the new cluster locations, the classification procedure is recursively executed for a specified number of iterations, or until the cluster locations become stable, and do not move with subsequent iterations.

Several problems can be identified with this procedure. First, the clusters are determined by statistical properties calculated from the entire image. Thus small spectral classes may be overlooked, and classes with overlapping spectral properties may be aggregated into a single class. As with other aspatial classifiers, the results tend to be noisy, with many isolated pixels of different classes, because the spatial properties are

ignored. Second, in common with other unsupervised classifications, the user has to assign the clusters to informational classes after the classification. This can be tedious and subjective if a large number of classes are identified.

The maximum likelihood classifier is a pixel-based supervised approach. Maximum likelihood classifies unknown pixel-based on multivariate probability density functions (pdf) of the classes of interest. However, the pdf of the classes are generally unknown. Therefore, the statistical properties of training data sets from ground reference data are used to represent the pdfs of the classes. Each unknown pixel is assigned to the class with the highest probability at the pixel location. The decision rule is as follows:

$$p(X|\mathbf{w}_c)p(\mathbf{w}_c) \geq p(X|\mathbf{w}_i)p(\mathbf{w}_i)$$

Where:

X : the spectral multivariate vector

$p(X|\mathbf{w}_c)$: pdf of X , given that X is a member of class c

$p(\mathbf{w}_c)$: *a priori* probability of class c in the image

i : class number among the m number of classes in the image

The *a priori* probability of a class is the probability of the class prior to classification, and can be approximated by the expected extent of the class in the image. However, in many cases the *a priori* probabilities of the classes are unknown, and are therefore simply set to an equal value for all classes. The pdf of the classes are calculated based on an assumption that they have a normal distribution. The probability of a pixel belonging to a particular class can then be calculated using the class variance and mean:

$$p(X|\mathbf{w}_c) = \frac{1}{2^p p^{p/2} |\Sigma|^{1/2}} \exp[-1/2 (X - \mathbf{m})^T \Sigma^{-1} (X - \mathbf{m})] \dots\dots(5-1)$$

Where:

- p : number of variables used
- Σ : variance-covariance matrix of class i
- $|\Sigma|$: determinant of variance-covariance matrix of class i
- \mathbf{m} : mean vector of class i

To simplify the calculation process, equation (5-1) is modified by multiplying the equation by $2\mathbf{p}^{p/2}$ and taking the natural log. Because $2\mathbf{p}^{p/2}$ is a constant, and the natural logs of any sequence of numbers does not change the order of their magnitudes, the resultant likelihoods (D) can be used as surrogates for probabilities.

$$D = [\ln(p(X|\mathbf{w}_c)p(\mathbf{v}_c))]2\mathbf{p}^{p/2} = \ln(p(\mathbf{v}_c)) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(\mathbf{X} - \mathbf{m})^T \Sigma^{-1}(\mathbf{X} - \mathbf{m})$$

The maximum likelihood classifier is an aspatial approach in which each pixel is independently classified. Pixels with low pdf values have a high probability of being misclassified. Figure 5-1 shows the pdfs of two spectral classes. Although the pdf of the two classes extend over all potential DN values, the pdfs have been arbitrarily terminated at a low probability level. The area marked with diagonal lines is the overlap region of for the class pdfs. The decision rule for this method is that all pixels are assigned to the class with the higher pdf for that spectral value. For example, even if a pixel with the value of “a” belongs in reality to class B, it will be classified as class A (Figure 5-1). This is an inevitable result of overlapping class pdfs.

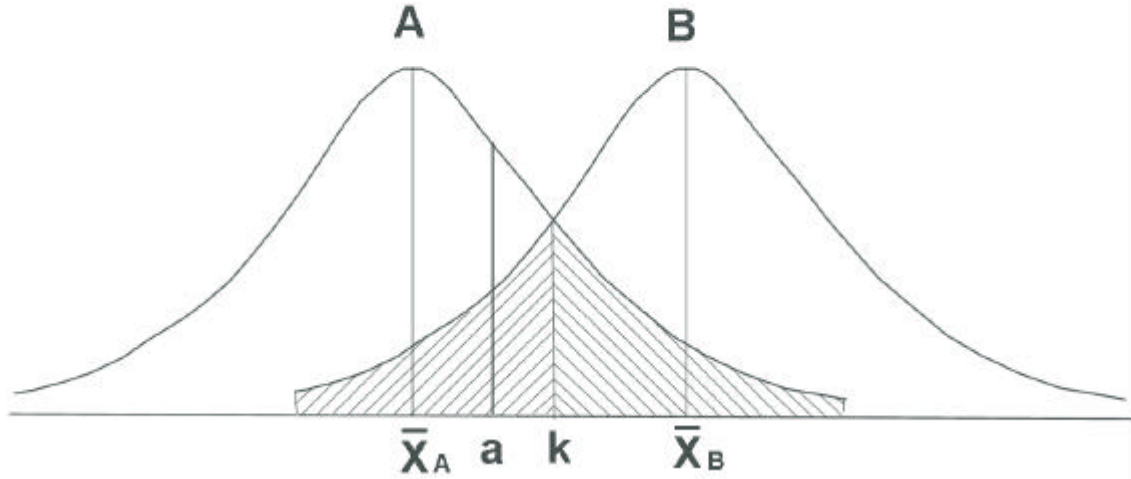


Figure 5-1. The decision rule of a pixel-based maximum likelihood classifier.

The ECHO classification (Kettig and Landgrebe, 1976) has been discussed in Chapter III with regards to segmentation. After segmentation, ECHO uses maximum likelihood classification on the means of the groups of pixels. Pixels that are not incorporated into groups, are classified using the standard maximum likelihood approach.

5-3. Methods

Instead of pixels, groups of pixels that form image segments were used for image classification in this study. There are few studies that evaluate the use statistic of segmented regions for classification (Kettig and Landberge, 1976; Meyer *et al.*, 1996; Gougeon, 1995a; Janssen and Molenaar, 1995). However, most studies employing aggregated information focus on first order statistics and only use second order statistics

to a limited extent. In this section, new methods that exploit multivariate statistics to improve the image classification are suggested.

Figure 5-2 represents a conceptual comparison between traditional classification and the methods developed in this study. An example of the pixel-based approach (Figure 5-2, left) is the traditional supervised maximum likelihood classification. Within a patch, pixels from the outliers of the class distribution are likely to be misclassified. Window-based approaches use arbitrary groupings and return the value of the window to the

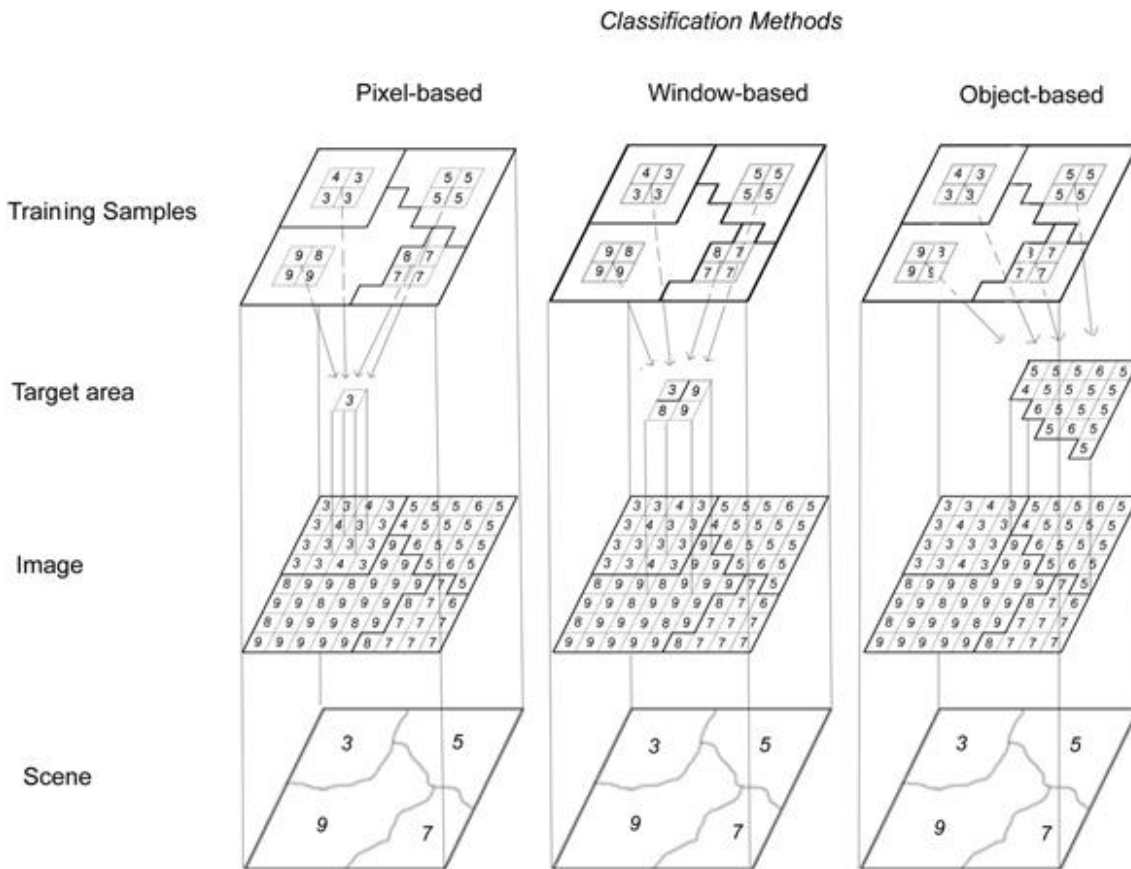


Figure 5-2. Comparison of object-based classification with traditional image classification approaches.

central pixel (Figure 5-2, middle). In the case of the object-based classification (Figure 5-2, right), patches are not expected to consist of pixels with completely homogeneous spectral radiances, but rather certain levels of variability are expected. This approach, therefore, incorporates a more realistic representation of real phenomena. The variation in an object is used as one characteristic of the object in this method, whereas it is an obstacle with traditional pixel-based classification methods. To treat this variation within objects, multivariate normal distributions were assumed for every group of pixels in each patch, and multivariate variance-covariance matrices were calculated. Three methods of exploiting this information were investigated: the divergence index, maximum likelihood based on the patch mean, and maximum likelihood with Gaussian pdf.

5-3-1. Classification with divergence index

Divergence is sometimes used as a band selection method, in which a subset of the original bands is chosen that best represent the overall data structure (Swain and Davis, 1978; Jensen, 1996). Band selection is often used as a method of reducing computing cost with the minimum effect on classification accuracy. Divergence is calculated from the average of the probability ratios between two classes:

$$L_{ij}(X) = \frac{p(X|\mathbf{w}_i)}{p(X|\mathbf{w}_j)}$$

Where:

$L_{ij}(X)$: likelihood between class i and j

$p(X|\mathbf{w}_j)$: pdf of class j , given that X is a member of class j

As shown for point k1 in Figure 5-3, the greater the probability that the point is class i , and the smaller the probability that it is class j , the greater the likelihood $L_{ij}(X)$ becomes. Furthermore, the more similar the probabilities of the two classes, the closer to 1 the likelihood $L_{ij}(X)$ becomes (Figure 5-3, point k3). Thus, the ratio will be increased by the combination of a large $p(X|w_i)$ and small $p(X|w_j)$. The more the two values have extreme values, the greater the increase is (Swain and Davis, 1978).

An efficient method of calculating the likelihood ratio is to take natural log for both side of the equation:

$$L'_{ij}(X) = \ln L_{ij}(X) = \ln \Pr(X | w_i) - \ln \Pr(X | w_j)$$

Divergence between class i and class j is defined by this logarithmic ratio as follows:

$$D_{ij} = E[L'_{ij}(X)|w_i] + E[L'_{ji}(X)|w_j]$$

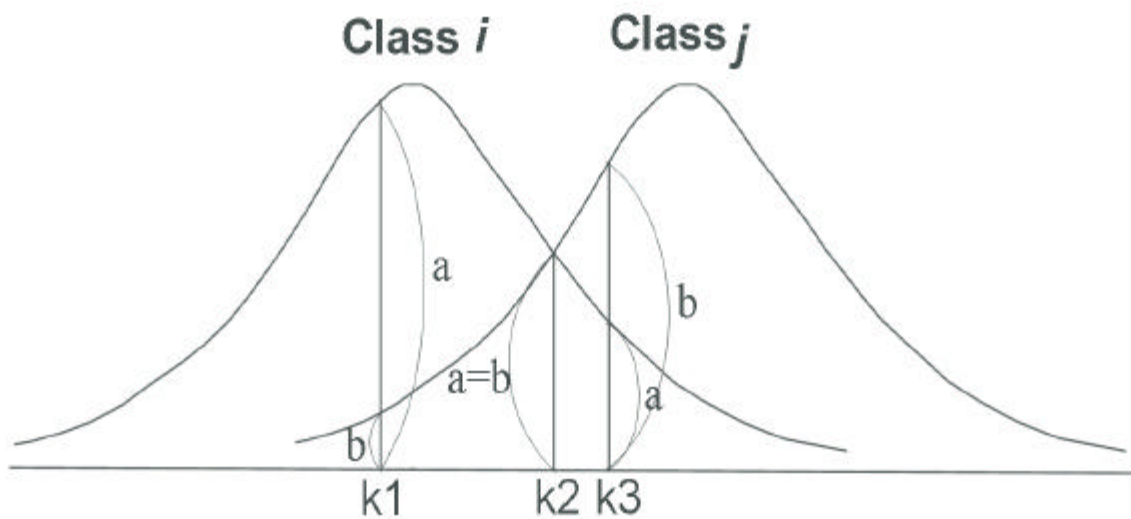


Figure 5-3. The decision rule for $L_{ij}(X)$. a is the probability for class i $p(X|w_i)$, and b is the probability for class j $p(X|w_j)$.

Where:

D_{ij} : divergence

$$E[L'_{ij}(X)|\mathbf{w}_i] = \int L'_{ij}(X)p(X|\mathbf{w}_i)dx$$

$$E[L'_{ji}(X)|\mathbf{w}_j] = \int L'_{ji}(X)p(X|\mathbf{w}_j)dx$$

When the two classes i and j are assumed to have a normal distribution, D_{ij} is represented with mean vectors and variance-covariance matrices as follows (Swain and Davis, 1978):

$$D_{ij} = \frac{1}{2} \text{tr}[(\Sigma_i - \Sigma_j)(\Sigma_j^{-1} - \Sigma_i^{-1})] + \frac{1}{2} \text{tr}[(\Sigma_i^{-1} - \Sigma_j^{-1})(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T]$$

D_{ij} is a function of $L_{ij}(X)$, and as mentioned previously, the greater the difference in the pdfs between two classes investigated, the greater D_{ij} becomes.

Divergence is useful in classification, because when the separability between two groups of pixels is small, the groups possibly belong to the same class. When divergence is employed for patch classification, D_{ij} is calculated for the patch and each of the training data sets. The training data set with the lowest D_{ij} values can be assumed to be the most similar spectrally to the patch. So the decision rule classification using this method is as follows:

$$D_{ic} \leq D_{ij}$$

Where:

D : divergence

i : patch number

- j : training data set, where j varies from 1 to the total number of training data sets
- c : class number for the patch

5-3-2. Maximum likelihood classification using the patch mean

Maximum likelihood classification with the patch mean uses a decision rule modified to use the mean vector of a group of pixels, instead of individual pixels. When the mean of the group is classified as belonging to a certain class, all the pixels in the group are assigned to that class. Therefore, the results are highly dependant on the accuracy of the segmentation employed. The decision rule is as follows:

$$p(\bar{X}|\mathbf{v}_c)p(\mathbf{w}_c) \geq p(\bar{X}|\mathbf{v}_i)p(\mathbf{w}_i)$$

Where:

\bar{X} : mean vector of a group

$p(\bar{X}|\mathbf{w}_c)$: probability associated with the mean of the group of pixels of class c ,
given that the mean vector \bar{X} is a member of class c

This method classifies each group of pixels as a unit. This will tend to minimize misclassification for isolated pixels with outlier spectral characteristics. ECHO employs this method for classification of the homogeneous regions identified through image segmentation (Landgrebe and Biehl, 2000). However, the variability of spectral measurements of spatially adjacent pixels within a group is collapsed into the mean vector of the group, and is not exploited in the classification.

5-3-3. Region-based maximum likelihood classification with pdf

The method suggested in this study can be summarized as a comparison of the pdf of an unknown group with pdfs of each of the training data sets. If two samples originate from the same population, the pdfs of the two groups should be similar to each other. Significantly, the distribution of radiance values that causes misclassification in pixel-based approaches (Swain and Davis, 1978), is critical information for the method developed in this study.

To simplify the explanation, suppose two normally distributed populations have means m_1 and m_2 , and standard deviation s_1 and s_2 , respectively. Figure 5-4 represents

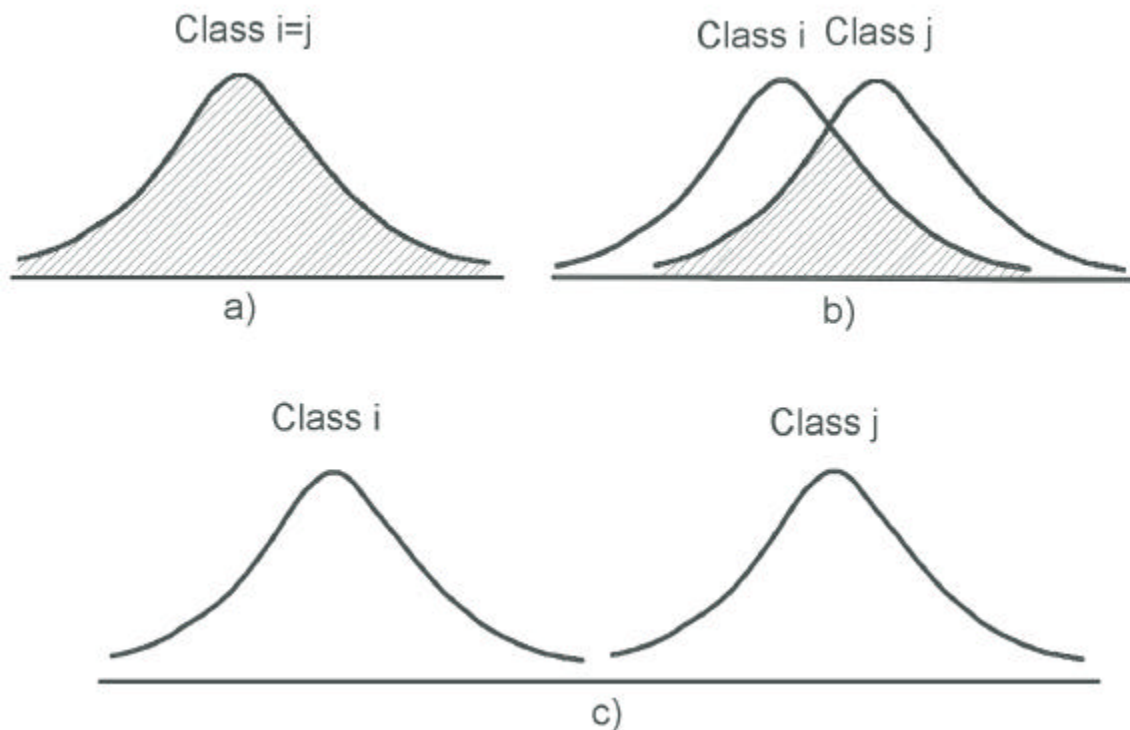


Figure 5-4. Likelihood measured with pdf. The areas with diagonal lines indicate the degree of similarity between two classes. (a) Two almost completely overlapping class. (b) Two partially overlapping classes. (c) Two almost completely separated classes.

three different cases that could occur. If two populations are very similar, then the two pdfs almost completely overlap (Figure 5-4a). If it is possible to estimate the area of the overlapped region, it should be close to 1, because the sum of all possible probabilities is equal to 1. However, if two populations are very different from each other, there should only be a very small overlap area for the two pdfs (Figure 5-4c). Thus it can be seen that the size of the overlapped area is proportionate to the similarity of the two pdfs. If the two pdfs are identical to each other, the overlapping area is equal to 1, if completely different, then 0, and the values between are an index of similarity (Figure 5-4b). The area of overlap can be found by integrating the relevant overlap portions of the two pdfs:

$$O_{12} = \int_{-\infty}^{+\infty} \left\{ \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{(X_m - \mu_m)^2}{2\sigma_m^2}\right] \right\} dX_m$$

Where:

O_{12} : likelihood index between X_1 and X_2

$m = 2$ for $X_1 \geq X_2$

$m = 1$ for $X_1 < X_2$

When the likelihood index is extended to a multivariate pdf, with p variables and multiple samples, the equation is modified as follows:

$$O_{ij} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left\{ \frac{1}{2\pi^{p/2} |\Sigma_m|^{1/2}} \exp\left[-\frac{1}{2} (X_m - \bar{X}_m)^T \Sigma_m^{-1} (X_m - \bar{X}_m)\right] \right\} dX_1 \dots dX_p$$

O_{ij} : likelihood index between X_i and X_j

i : patch id under investigation

j : training data set id under investigation

$m = j$ for $X_i \geq X_j$

$$m = i \text{ for } X_i < X_j$$

The decision rule in this study is extracted from the relationship between the likelihood and similarity as follows:

$$O_{ic} \geq O_{ij}$$

Thus a patch is assigned to class c if the maximum of the likelihood index values is found for the pdf comparison of the patch and training data set c . With this method, a very stable similarity index is obtained because the variance and covariance information, as well as the class mean, are all directly used. One disadvantage of this method is that the computing cost is high, specifically computing time.

5-4. Application of the pixel and group-based classification methods

The analysis procedure in this study comprises three stages (Figure 5-5). It is assumed that patches have previously been identified by image segmentation using the region growing process incorporating thresholding and region growing. In the first stage, statistics for the patches are computed. The statistics used were the same as those used in the region growing stage, including the mean vectors and variance-covariance matrices.

For the second stage, representative patches were selected to build a training data set for seven classes: Building, Road, Forest, Lawn, Shadowed Vegetation, Water, and Shadow. The patches selected as training data were treated as independent spectral classes within each informational class. This means that the selected patches were not aggregated into composite statistics for the seven classes. The likelihood index for each

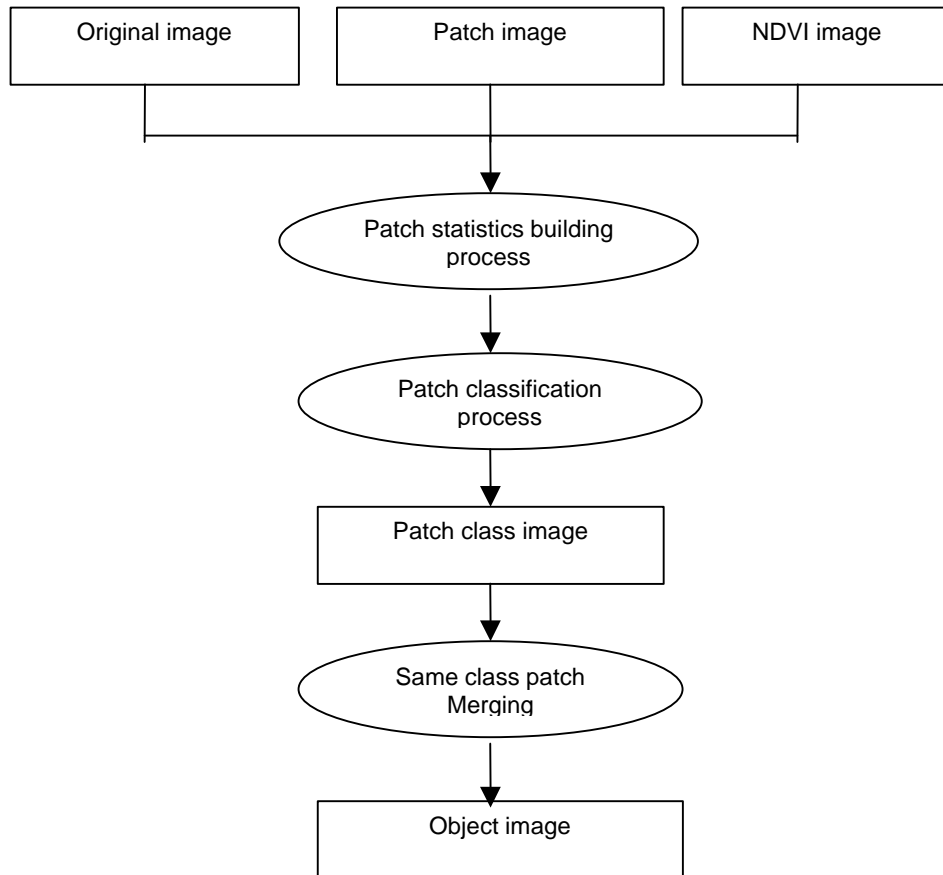


Figure 5-5. Outline of the spectral classification procedure using previously segmented data.

patch was computed for all individual training data set patches by the divergence index, the maximum likelihood using the patch mean, and the patch pdf.

Small patches with fewer than six pixels were excluded from the region-based maximum likelihood analysis, and treated as part of the “melt pond,” to use McDevitt and Peddada’s (1998) term. There were two reasons for identifying melting pond pixels. Firstly, because five variables were used in this study, patches with fewer than six pixels had less than the minimum number of pixels potentially required to characterize the

multivariate statistics. Secondly, the melting pond was assumed to represent objects that are not of direct interest, but rather extraneous objects such as cars, or chimneys on buildings.

For maximum likelihood with patch pdf, the range over which the pdf was calculated was limited to three standard deviations. The pdf is very low outside of this range, and is not expected to have much significance in the calculation. Excluding pdf values greater than three standard deviations has the advantage of reducing the computing cost. Figure 5-6 shows a one dimensional representation of the process. Within the pdf overlap region, the decision range was divided into ten equal cells. The probability of the center of each cell calculated for both the training and the patch classes, and the lower of the two probabilities is used for the cell height. After multiplying cell height by the width, the cell area is calculated. The total area of the overlap is then estimated by summing the cell areas (Figure 5-6). This procedure is modified for the multivariate case by dividing the multidimensional overlap region into 10^n cells, where n is the number of bands. For two bands a volume of the overlap region is calculated, and for three or more bands a hypervolume is calculated. For this work, five bands were used, thus, 10^5 cells were calculated for each likelihood index. The patch was assigned to the class with the highest likelihood after the unknown patch is compared with each patch in the training data set. In the next step of the classification, melting pond pixels are classified. These small patches are treated as noise, and therefore assigned to an adjacent class. If the patch is surrounded by a single class, it is assigned to that class. In the general case, however, the patch is adjacent to more than one class. In this case, the patch is assigned to the adjacent class with the most similar DN values in the green band (Band 2). A more sophisticated,

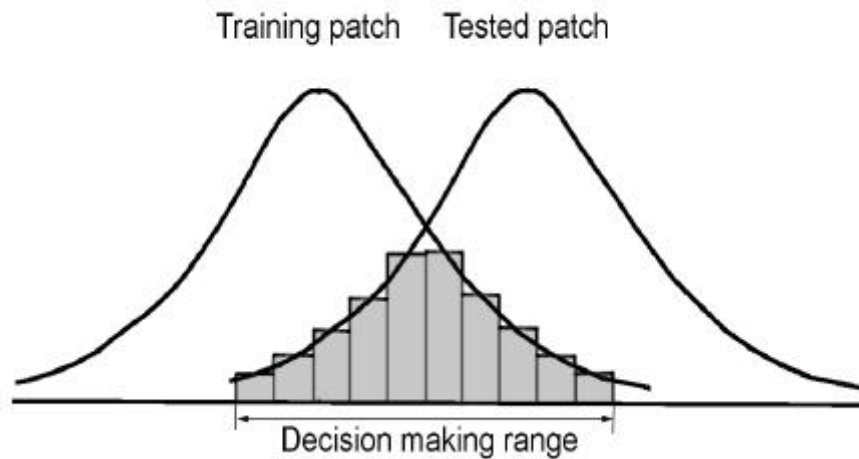


Figure 5-6. Maximum likelihood calculation utilizing patch pdfs.

multivariate approach was not used because of the small sample size of these patches. In the final step, adjacent patches of the same class were merged to form objects.

ERDAS Imagine was used to conduct the traditional pixel-based classifications. The unsupervised ISODATA program (Tou and Gonzalez, 1974; ERDAS, 1999) was executed with 24 clusters. After classification, the 24 clusters were assigned empirically to the most appropriate class among the seven classes based on the ground truth and knowledge of the area.

For each of the supervised classification methods, the same training data sets used. Multispec version 1.2, IBM-compatible (Landgrebe and Biehl, 2000) was employed for the ECHO classification.

5-5. Evaluation of the classifications

Figure 5-7 shows the results from the six previously mentioned methods. (Note, for

ease of access, Figure 5-7a-g is incorporated at the end of this chapter.) To compare the accuracy of the six methods, error matrices for the kappa index and errors of commission and omission were produced using the IDIRSI program ERRMAT (Eastman, 1997) (Table 5-1). Ground reference maps for the accuracy evaluation were produced using photo-interpretation and expert knowledge for three parts of the study area: Downtown Morgantown, a medium density residential area, and a forested stream valley.

The overall kappa values of the unsupervised and supervised pixel-based classifications were 0.610 and 0.687, respectively. The field-based ECHO classifier produced a 0.605 kappa, a value that is lower than that of the pixel-based classifiers. The lowest accuracy, 0.254, was obtained with the classifier that utilized the divergence index. The maximum likelihood classifier using the patch mean resulted in a relatively high kappa value of 0.735. Maximum likelihood classifier with pdf produced the overall best accuracy, 0.783.

Table 5-1. Summary accuracy statistics for 7 classes by the 6 classification methods used in this study.

		Building	Road	Forest	Lawn	Shadowed Vegetation	Water	Shadow	Kappa index
ISODATA	CERR	0.363	0.489	0.135	0.440	0.567	0.024	0.062	0.610
	OERR	0.487	0.320	0.147	0.046	0.373	0.981	0.352	
MHL with pixel	CERR	0.391	0.347	0.063	0.366	0.440	0.123	0.137	0.687
	OERR	0.202	0.412	0.133	0.214	0.297	0.308	0.359	
ECHO	CERR	0.301	0.569	0.077	0.491	0.240	0.084	0.251	0.605
	OERR	0.383	0.304	0.191	0.684	0.497	0.150	0.320	
Divergence index	CERR	0.663	0.643	0.199	0.545	0.834	0.654	0.808	0.254
	OERR	0.334	0.468	0.733	0.552	0.925	0.106	0.911	
MLH with patch mean	CERR	0.309	0.291	0.104	0.304	0.360	0.095	0.048	0.735
	OERR	0.194	0.292	0.042	0.503	0.691	0.232	0.398	
MLH with patch pdf	CERR	0.225	0.243	0.101	0.294	0.306	0.042	0.054	0.783
	OERR	0.170	0.249	0.045	0.480	0.603	0.089	0.234	

Looking at the results in more detail, the unsupervised classifier resulted in many isolated pixels and small clusters, as expected (Figure 5-7 b). The Water class in the region of the stream was almost completely misclassified as Building with this method. The stream has exposed and shallow covered rock that is apparently spectrally similar to the materials from which buildings are constructed. Building was also misclassified as Road, and consequently the Building omission error was relatively high (Table 5-1). Pixel-based supervised classification (Figure 5-7 c), like the unsupervised classification, resulted in a rather noisy classification. The classes of Buildings and Roads were extensively confused, resulting in high errors of commission and omission for both classes. However, compared to the unsupervised classification, the confusion between Building and Water was dramatically reduced for the pixel-based maximum likelihood classification.

The ECHO classification (Figure 5-7 d) resulted in significant errors of commission, especially between the classes Road and Lawn, and high omission error between Shadowed Vegetation and Lawn. The ECHO classifier uses the same decision rule as the maximum likelihood classifier utilizing patch means, but ECHO employs a different segmentation rule (Kettig and Landgrebe, 1976). In comparing the segmentation of ECHO to the that of the ridge finding process (Figures 5-7 e, f, g), it is apparent that ECHO produces a far more noisy segmentation. In particular, the shape of the buildings is not particularly clear. ECHO's segmentation is designed to allow isolated pixels that are not similar to their neighbors to be classified independently. Therefore, although ECHO is likely to produce a less noisy classification than the pixel-based classifiers, the results can still be poor in places.

The divergence approach (Figure 5-7 e) had greatest problems with Forest vegetation, Shadowed Vegetation, Shadows and Roads. In addition, misclassified water is scattered throughout the downtown and residential areas. It is apparent that the divergence approach was the least successful method of incorporating spatial information in the classification process.

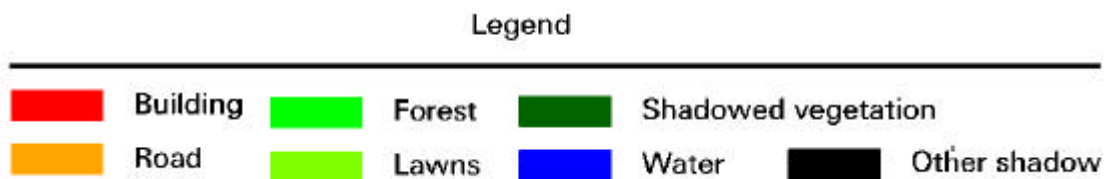
The maximum likelihood classifier using the patch mean (Figure 5-7 f) yielded a visually pleasing classification, and the second best overall accuracy. The higher classification accuracy of the maximum likelihood classification with patch pdf is most likely a result of the incorporation of differences in the kurtosis of classes through the variance-covariance matrix data. When only the patch mean is used in the classification, such differences are suppressed. The particular classes that were less well classified in the maximum likelihood using the patch mean, compared to the patch pdf, were the Building and Road classes. But the computing cost for classification with the mean was much lower than with the pdf. Thus, the classifier with the patch mean is an efficient alternative to classification with pdf.

The maximum likelihood classification with pdf produced higher accuracy than any other classifier (Table 5-1). The segmentation suppresses isolated pixels and small clusters (Figure 5-7 g), and thus classification error resulting from high within object variance was efficiently controlled by this method. However, a number of cases of confusion arose between Building and Road, and Lawn and Forest. The confusion between Lawn and Forest can be related to segmentation. Although these two classes generally had sufficient spectral difference between them for good classification, in some cases the low contrast boundaries between Lawns and Forest areas resulted in these

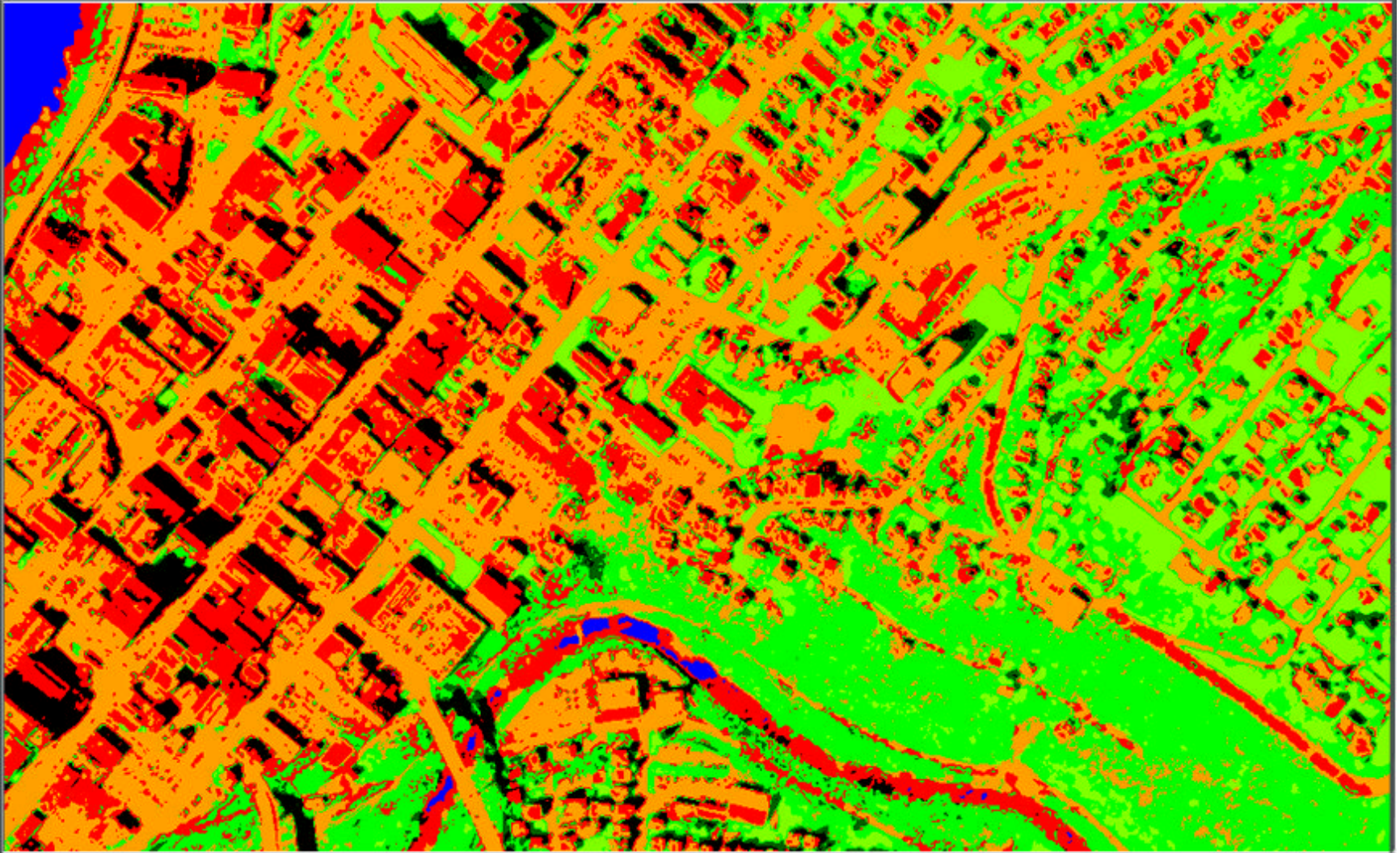
regions being merged into a single patch. The confusion between Building and Road was not a result of segmentation as generally these two classes were well delineated. However, confusion occurred because the spectral radiances of the two classes were sometimes very similar. This arises because materials such as asphalt, stone and concrete are used for both building roofs and roads.

As part of the classifications carried out using maximum likelihood, all pixels were assigned to the class with the highest likelihood. This is a relative, not an absolute measure. Thus even classes that result in very low likelihood when compared to all the training data sets are classified. It is possible that a region is not represented by any of the training data sets, and this should be identified. In future work, it may be desirable to establish an absolute minimum maximum likelihood for classification. Patches that fail to meet the minimum value would be flagged as unknown.

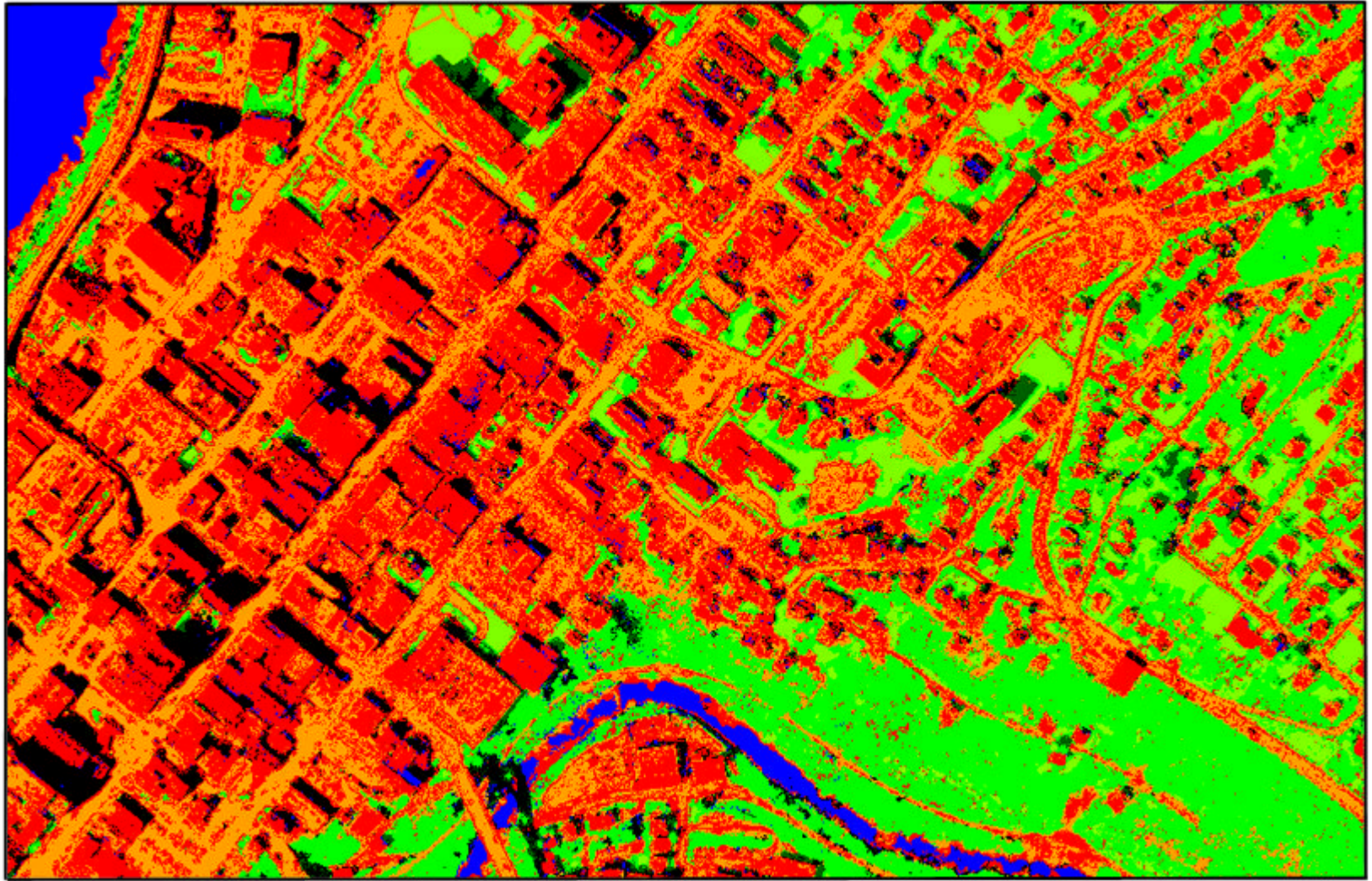
Figure 5-7. Results of the classifications. a): Legend. b): ISODATA from ERDAS Imagine. c): Maximum likelihood classification from ERDAS IMAGINE. d): ECHO from Multispec. e): Divergence index. f): Maximum likelihood classifier with patch mean. g): Maximum likelihood classifier with patch pdf.



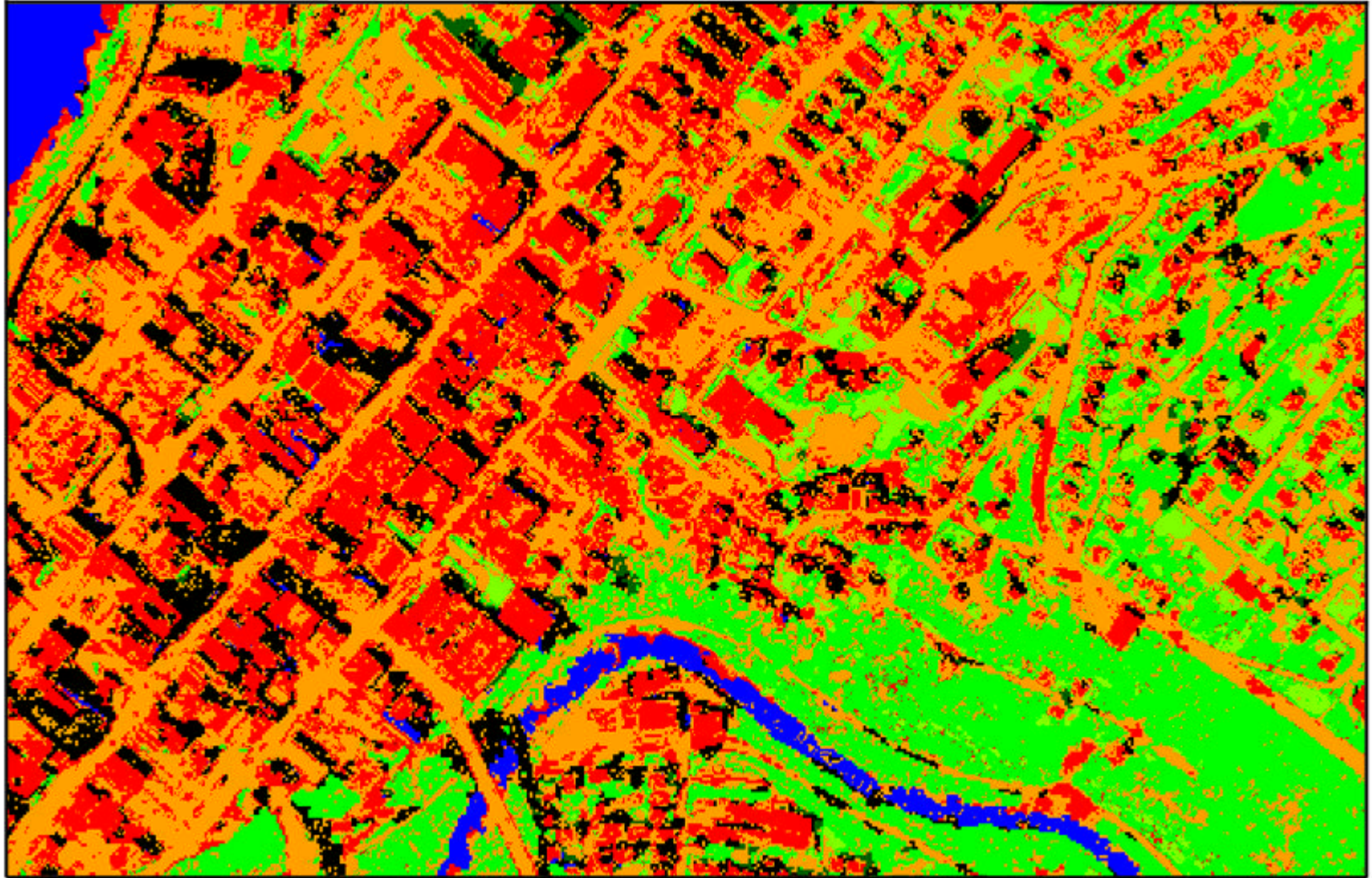
a)



b)



c)



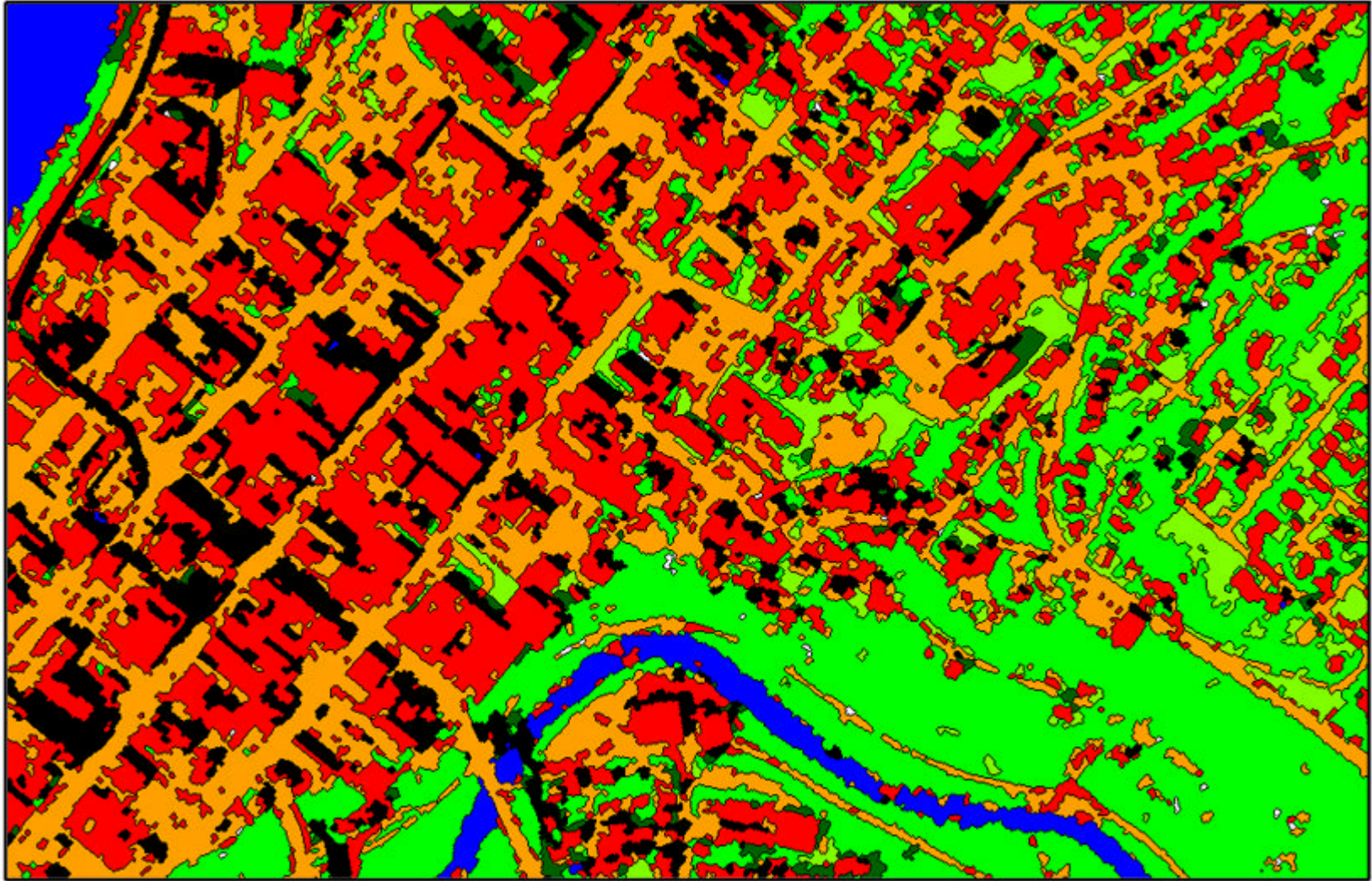
d)



e)



f)



g)

Chapter VI

OBJECT-BASED LAND COVER CLASSIFICATION

In the previous chapters the procedure for identification of objects in remotely sensed imagery has been presented. In many cases it is desirable to present this data in conjunction with aggregated land cover or land use information at a coarser spatial scale. Therefore, as the final step in the image analysis procedure, a land cover map was synthesized from the image object map. In this study, the term land cover class is used in a broad sense relating to building density and related measures that can be used to infer land use. One advantage of producing land cover information in this manner is that a more detailed definition for land cover type is possible than using traditional methods with low spatial resolution data. With traditional methods it is necessary to infer land cover from the average spectral properties of the classes. However, with high-resolution data it is possible to make direct estimates of object properties such as building density.

The land cover analysis procedure has three stages (Figure 6-1). The study area was segmented into regions of relatively homogeneous urbanization levels. The index for this segmentation is the “Urbanized Rate,” defined as the local proportion of Building and

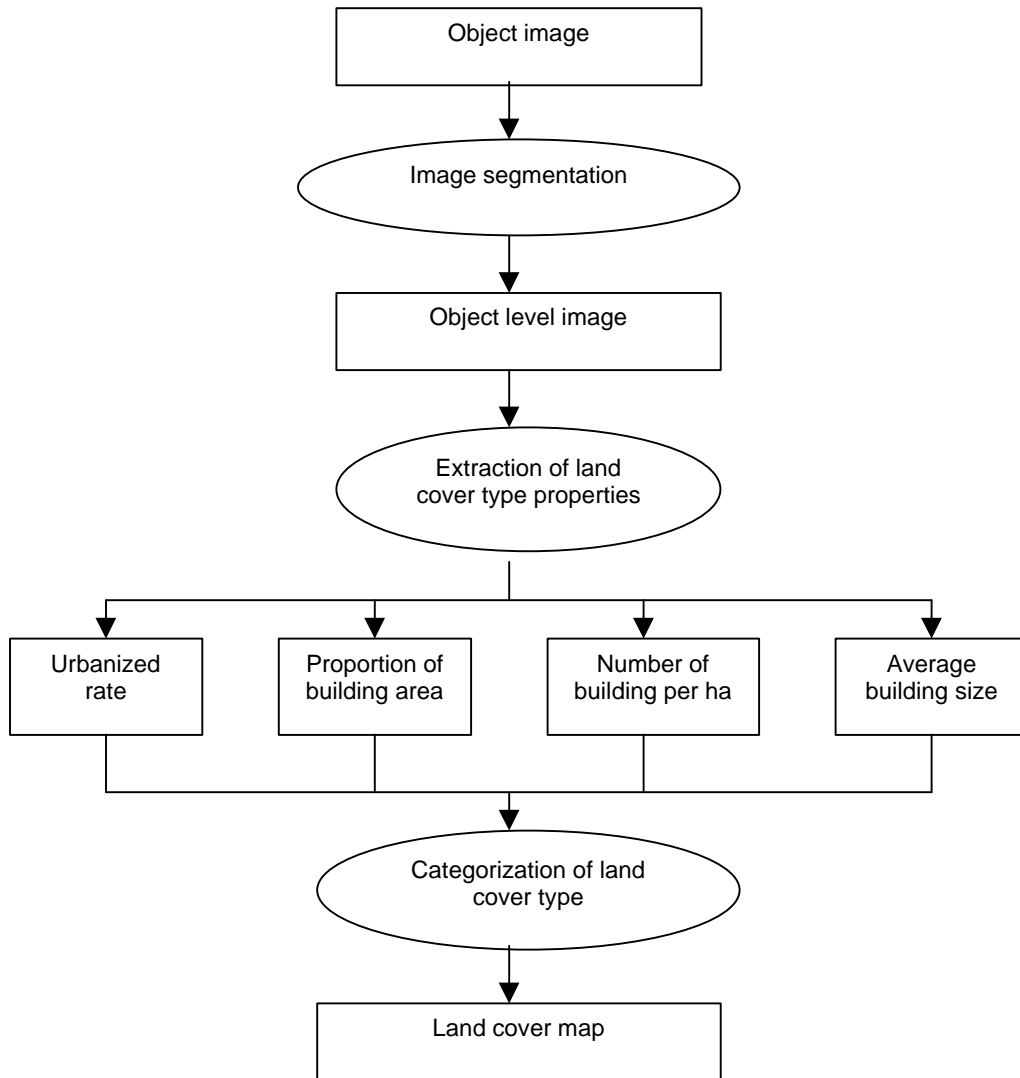


Figure 6-1. Land cover type map classification

Road cover in an area. In calculating the rates, shadowed areas were excluded so as not to bias the results. Road and Building were used for calculating the Urbanized Rate because of their potential value in differentiating the major land cover classes. For example, the USGS land use and land cover classification system (Anderson *et al.*, 1976) notes that level II commercial and service areas are associated with large buildings such

as office complexes, warehouses, and retail establishments, as well as driveways, parking lots, waste disposal areas, and landscaped zones. On the other hand, Anderson *et al.* (1976) define low density residential as an example of a Level III class in which the lot for each house is greater than one acre. The low density residential class is therefore characterized by a much lower urbanized rate than commercial and service areas. Thus land cover attributes such as the Urbanized Rate provide a quantifiable metric that can be used to infer land cover, and possibly even land use.

The Urbanized Rate was computed from the object classification map using a moving window, 100 by 100 pixels in size. This window size was chosen based on Hodgson's (1998) study and the definition of the low density residential class in Anderson *et al.* (1976). Hodgson (1998) examined classification accuracies with window sizes that ranged from 15 by 15 meters to 150 by 150 meters, and reported that cognitive classification accuracies improved with larger window sizes, up to approximately 60 by 60 meters in size. Because a 70 by 70 meter window is needed to cover one acre, a window size greater than this was needed to evaluate low density residential areas, where densities are lower than one house per acre. Consequently if the window size is smaller than this, the Urbanized Rate will produce unwanted spatial variation.

The edge detection and ridge method was applied to the Urbanized Rate image to find homogeneous areas. Edge enhancement was carried out using the 11 by 11 linear shape operator. The segments generated were refined first by thresholding with a level of 15. This initially produced many small segments, which were eliminated through two passes with an erosion method. The intended minimum segment size was 50,000 m². If the minimum segment size is set too small, the variation within a relative homogeneous

region might dominate the result. The thresholds for the maximum polygon size eroded for the two passes were 100 and 700 pixels, respectively. The two step process, using an increased size of threshold on the second pass, was required because otherwise too many polygons were eliminated in the first pass, leading to displaced boundaries. In the next step, for each segment, an average Urbanized Rate was computed, and then categorized into four classes: 1-19, 20-39, 40-89, and greater than 90 per cent. Then the same class regions were merged. The results still include a number of regions smaller than the intended minimum size. These regions were merged into adjacent regions by further erosion. This had the effect of smoothing the data, and removing distracting minor variation associated with small regions. The results are shown in Figure 6-2a.

Within each region identified as having a relatively uniform Urbanized Rate, additional land cover parameters were calculated: the percentage of Building area, the number of building polygons per hectare, and average building polygon size for each region. Table 6-1 and Figure 6-2b, c and d show the results of this step. The percentage of Building area ranged from 4 % to 45 %. As expected, regions with a high Urbanized Rate tend to have a high percentage of Building area. the average size of buildings was

Table 6-1. Land cover properties by region

Region	Urbanized Rate (percent)	Building as percentage of total area	Number of Buildings / ha	Average building size(m ²)
1	16	5	2	163
2	89	45	6	255
3	61	30	11	199
4	35	16	10	128
5	19	6	4	124
6	66	35	12	195
7	11	4	4	88

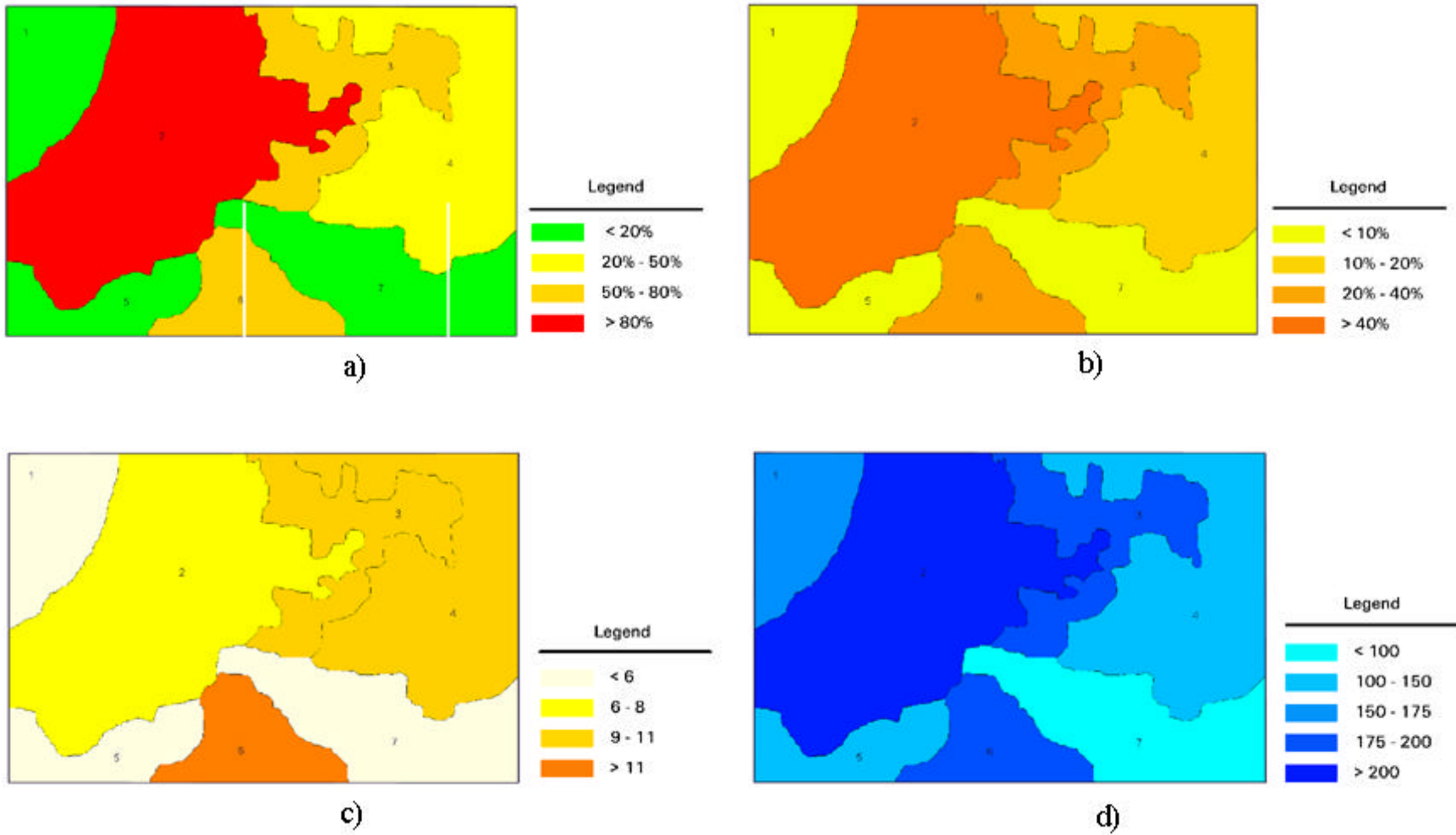


Figure 6-2. Land cover properties of regions. (a) Urbanized rate, (b) the percentage of building areas, (c) number of building polygons per hectare, (d) average building polygon size

also relatively large for regions with a high Urbanized Rate. The number of buildings per hectare map produced a somewhat different pattern from other indicators, particularly for the large numbers of buildings in the intermediate Urbanized Rate region.

Based on the properties of land cover, the regions were generally categorized into four classes: High Urbanized Rate and Large Building Size Region, Intermediate Urbanized Rate and Intermediate Building Size Region, Low Urbanized Rate and Small Building Region, Water/Water Front Region, and Forest Region (Figure 6-3).

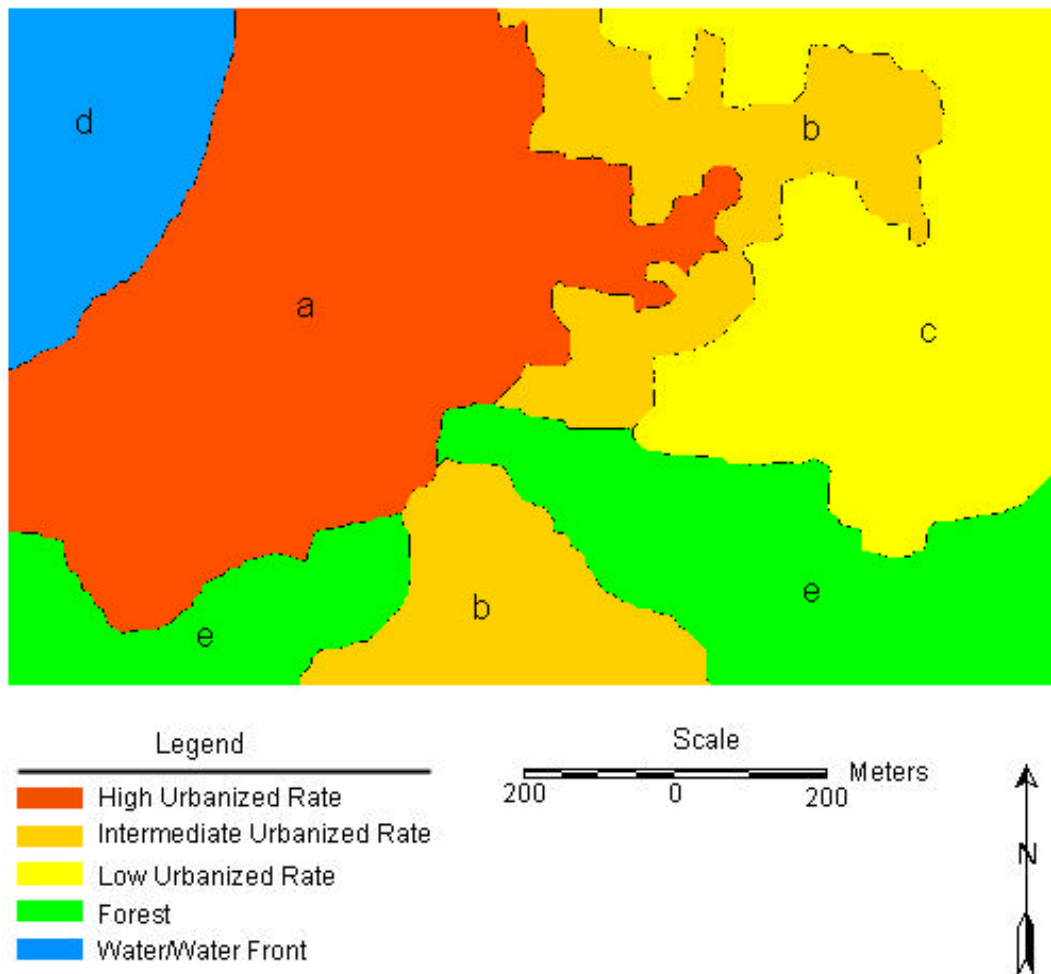


Figure 6-3. Land cover classification map.

Region **a** in Figure 6-3, the High Urbanized Rate and large building size region, comprises most of downtown Morgantown. This region is highly urbanized, and large buildings are common in this region. The Urbanized Rate of this region was over 80 percent, and average building polygon size was over 250 m². The average building polygon size was over-estimated in this region. This occurred because the distances between buildings were small, the background between buildings was not well detected at the given spatial resolution, and the buildings and background were spectrally similar. Region **b**, the Intermediate Urbanized Rate and intermediate building size region, is adjacent to region **a**, and is transitional from the highly developed region **a** to the residential class. The average building polygon size of this region was intermediate, at around 200 m², and included both large and small buildings. The Urbanized Rate of this region was about 60 percent. The southernmost of the three region **b** zones is an old residential area in this city. The building polygon size of this area was also overestimated, because the houses are built very close together, and therefore many houses on the same block were merged into the same object. Region **c**, the Low Urbanized Rate and small building region, has a low Urbanized Rate that was around 35 percent, but the number of building polygons per hectare is one of the highest regions. This region is a comparatively low density residential area, with relatively small houses separated from each other by lawns and gardens. Region **d** and region **e** are Water/Water Front and Forest Regions, respectively. The Urbanized rate of these regions were less than 20 percent. The Forest Regions included only a small number of buildings on the Decker's Creek floodplain and adjacent steep slopes, less than five buildings per hectare.

The overall land cover pattern of the study area shows the expected pattern of a highly developed core, with adjacent transitional zones, which are in turn surrounded by low density housing. Thus development decreases in intensity with distance from downtown. The presence of the Forest and Water classes is determined by the topographic constraints, and does not disrupt the overall pattern.

The final result of object-based classification (Figure 6-3) was vectorized to examine the possibility of incorporation with GIS. The software used was ERDAS IMAGINE, which supports vectorization of thematic raster maps using ArcInfo functionality. The result is shown in Figure 6-4, which shows the potential of incorporating remotely sensed

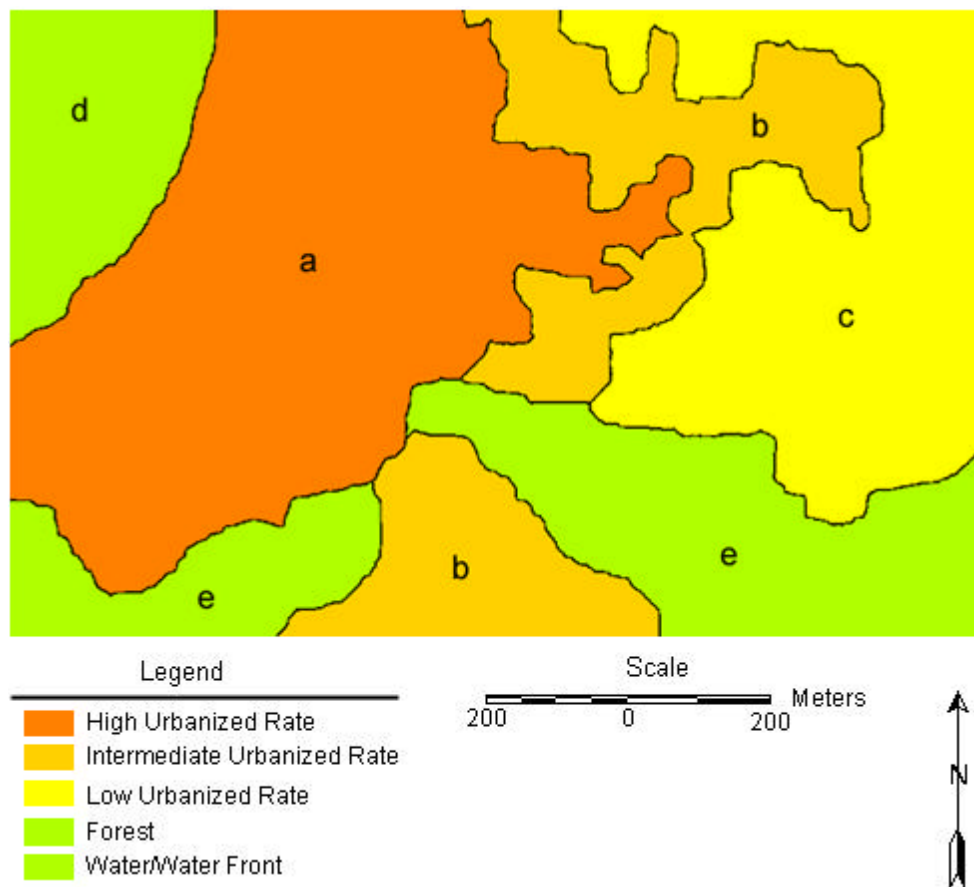


Figure 6-4. The vectorized land cover map.

data and GIS. Specifically, the problem of isolated class pixels and small cluster classes has been overcome. On the other hand, the fundamental difference of data structure related to the square shapes of the pixels creates a rather blocky vectorization that could possibly be overcome by additional smoothing.

Chapter VII

CONCLUSIONS

This study produced a new feature-based image segmentation and classification approach specifically designed for high spatial resolution imagery. An important part of this work was the identification of image objects through the development of a reliable method of image segmentation. The new classification method resulted in improved results at both the image object scale and a richer attribution at the aggregate land cover scale.

This research made a contribution to the growing field of analysis of high spatial resolution imagery. Although much of the latter research has in the past focused on trees and forested areas, high spatial resolution data could be a very efficient source of information for urban land use spatial information systems. There is a need for highly detailed information, for example in urban planning and cadastral mapping. Automated methods have the potential for more rapid and more consistent mapping than human interpretation.

The results of this research should be especially useful to under-developed or developing countries, especially those that do not have sufficient large scale maps of land use/land cover for urban and regional planning. This approach could also be used for updating large scale maps, because land use/land cover is continuously changing. However, new procedures would need to be written to draw on the existing data if the approach were used for updating maps.

The methods developed in this research are important not just because they produce more accurate results that show the spatial patterns more clearly because of their lack of distracting high frequency noise. The delineation and attribution of image objects, rather than classified pixels, is an important step toward integrating remote sensing with GIS. Attributed image objects are conceptually much closer to the vector data structure of most commercial GIS packages. Furthermore, the image processing procedure used in this study is based on an image-database concept, which could be used to incorporate image information into the metadata and attribute tables in the GIS. For example, information on classification confidence for each object could be tracked, as well as details such as the radiance properties of the sub-patches that made up the object. The land cover map is also unlike a conventional image classification. The land cover map is based on segmented regions, and has attribution, such as the number of building polygons per unit area, information that is not normally available directly from remote sensing classifications.

Each step in the image analysis sequence incorporated new approaches to feature extraction and attribution. Starting with the edge enhancement, linear shaped operators were found to be more effective at identifying curved edges than traditional square

shaped operators. Furthermore, linear operators are much less sensitive to scale issues than square operators, the making practical use of linear operators far more robust. This is important because high spatial resolution appears to require large window sizes to produce stable measures of spatial characteristics.

The ridge finding method draws on the edge enhancement data in order to identify closed polygons with a greater value than an arbitrary threshold. The ridges that had a width of more than one pixel were thinned to a one-pixel width ridge by eroding lower edge enhancement measurement pixels. This approach represents an improvement over conventional image erosion methods, which tend to isolate the center of the eroded features. In this study, erosion was applied to identify the highest ridge values, irrespective of their positions on the ridge cross-section. This will tend to result in a more accurate estimation of edge location, although it has the disadvantage of being more computationally intensive.

The region growing method of image segmentation used multivariate statistics for identifying adjacent sub-patches of similar radiance properties. Previous region growing research focused mainly on first order statistics such as the mean (Fjørtoft *et al.*, 1998), Euclidean distance (Mason *et al.*, 1988), and occasionally the variance of segments (Moscheni *et al.*, 1998), but not the interaction effects with other variables. In this study, second order multivariate statistics were used.

The classification approach used in this work is based on a polygon concept for remotely sensed data. This method is characterized by the identification of objects, the attributes of which are used in a multivariate analysis. The object-based approach resulted in a pleasing simplicity of spatial structure compared to the noisy patterns of

traditional pixel-based classification. The new method does not require objects to have a homogeneous radiance. Instead, a hierarchy of sub-patches and patches are identified prior to classification, which produces image objects. This procedure allowed a complex, many-to-one relationship between spectral classes and informational classes, even for one object. The pdf's based on the multivariate variance-covariance matrices of the spectral measurements of individual objects were directly compared with training data sets that were used in the classification. This improved the classification accuracy from 0.610 and 0.687, the kappa values respectively from ISODATA and a maximum likelihood classifier, to 0.785 by maximum likelihood with patch pdf. Even when only the patch means were used with the maximum likelihood classifier, the classification accuracy was 0.736, the second-best value obtained. The patch-based classifier using the divergence index produced the lowest accuracy.

Land cover maps were produced by aggregating objects to a coarser scale. With traditional pixel based approaches land cover is inferred from spectral classes. The method developed in this study dealt used the Urbanized Rate Index, an approach based on the proportion of Road and Building polygons in local regions. Following segmentation based on the Urbanized Rate, related statistics such as average sizes of building polygons and number of building polygons per unit area were employed for land cover classification.

The raster polygon concept central to this research has broad significance for remote sensing image processing and classification. For example, the raster polygon approach has the potential for extension to include any number of additional attributes, such as linearity, orientation, and fractal dimension. In addition, context, such as the proportion

of area surrounding a building that is covered by lawn, or location of shadow relative to the building, could be also stored in the database, and used in subsequent processing. Linearity and orientation could also be used in road identification. Thus raster polygons have great potential for image processing.

The methods developed in this research were applied to a Morgantown, WV study site. The study site has a wide range of cover types and topographic settings, as well as a complex road network. Consequently, this area represents a challenging test site. If the methods were applied to other areas, new methods would not necessarily be required. The high degree of success of the classification for this complex site suggests the methods are sufficiently generic to apply over differing geographic areas and a variety of urban densities. The overall accuracies for alternative areas will, of course, be scene and context dependent. Certain scenes are inherently simpler to classify, and therefore should produce better results. For example, buildings that are isolated from each other by land cover that contrasts strongly with the buildings should produce better results than scenes where the spectral properties of building roofs match those of the ground between them.

This research represents a first step in developing an analytical procedure specially adapted to high resolution imagery. Future work should consider the following issues:

- The edge enhancement method used in this work was based on a single band. However, a multivariate approach to edge detection is likely to provide a more sophisticated measure of land cover edges.
- The thresholding procedure following the ridge finding process was applied on a per-pixel basis. If just one pixel on the edge of a polygon is removed, the rest of the edge will be suppressed in the cleaning process. This may cause real edges

to be lost, although in this study a low threshold was chosen to minimize the chances of this occurring. More complex rules for thresholding may allow a higher threshold value to be chosen, eliminating some spurious edges and at the same time possibly retaining more real edges. For example, applying the threshold to the mean or median edge enhancement value of a polygon edge may be preferable. Alternatively, if a per-pixel basis is used for thresholding, a fuzzy tolerance may be applied so that edges with gaps that are small with respect to their overall length, are not eliminated, but instead the suppressed edge pixels are reincorporated into the edge image.

- This study focused on analyzing spectral and textural information rather than the geometric structure and associated information. However, there is a limitation in identifying objects based purely on spectral and/or textural information. Man-made objects tend to have a distinctive structure. In order to incorporate such information in automated feature extraction, higher spatial resolution than the one meter imagery used in this study, might be needed. For example, the data used for this study was insufficiently detailed for analysis of buildings that are in close proximity. Thus in many instances in the central business district of Morgantown, buildings within a city block were aggregated into a single composite image object.
- An important issue in automated feature extraction arises from variations in radiance due to slope/aspect differences, for example the opposing sides of a pitched roof. Furthermore, shadows are not only lower radiance regions, but have a different spectral illumination than sunlit areas. Normalizing for such

differences is ideally achieved using a digital elevation model, however, developing such a high resolution digital elevation model remains very expensive, at least at this stage. Therefore, empirical, scene-based normalization approaches need to be developed, to reduce spurious segmentation associated with shadows and slope/aspect changes.

In conclusion, this study has shown that high spatial resolution imagery holds great promise for automated feature identification. In addition, it has been shown that automated feature extraction with high resolution imagery is more effective when new methods are used that draw on the spatial information present in such imagery.

APPENDIX

1. Error matrix for the result of unsupervised classification with ISODATA.

Error Matrix Analysis of EVALU (columns : truth) against UNSP7R1 (rows : mapped)

Class	1	2	3	4	5	6	7	Total	ErrorC
1	8824	1175	43	0	10	1838	1965	13855	0.3631
2	7497	10037	1593	6	0	15	497	19645	0.4891
3	713	2778	28365	202	90	105	541	32794	0.1351
4	107	711	2690	4467	2	0	0	7977	0.4400
5	11	31	529	5	475	0	46	1097	0.5670
6	1	0	0	0	0	41	0	42	0.0238
7	61	19	12	0	181	99	5614	5986	0.0621
Total	17214	14751	33232	4680	758	2098	8663	81396	
ErrorO	0.4874	0.3196	0.1465	0.0455	0.3734	0.9805	0.3520		0.2896

ErrorO = Errors of Omission (expressed as proportions)

ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0026 (0.2870 - 0.2922)

95% Confidence Interval = 0.0031 (0.2865 - 0.2927)

99% Confidence Interval = 0.0041 (0.2855 - 0.2937)

KAPPA INDEX OF AGREEMENT (KIA)

Overall Kappa = 0.6098

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed vegetation, 6: Water, 7: Shadow

2. Error matrix for the result of MHL classification with pixel.

Error Matrix Analysis of EVALU (columns : truth) against SPE2501 (rows : mapped)

	1	2	3	4	5	6	7	Total	ErrorC
1	13721	4689	979	175	13	481	2460	22518	0.3907
2	3028	8670	1073	12	40	44	405	13272	0.3467
3	154	849	28810	799	96	2	49	30759	0.0634
4	31	328	1762	3681	0	0	0	5802	0.3656
5	5	32	288	13	533	0	81	952	0.4401
6	43	20	29	0	1	1452	111	1656	0.1232
7	232	163	291	0	75	119	5557	6437	0.1367
Total	17214	14751	33232	4680	758	2098	8663	81396	
ErrorO	0.2029	0.4122	0.1331	0.2135	0.2968	0.3079	0.3585		0.2331

ErrorO = Errors of Omission (expressed as proportions)
 ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0024 (0.2306 - 0.2355)
 95% Confidence Interval = 0.0029 (0.2302 - 0.2360)
 99% Confidence Interval = 0.0038 (0.2293 - 0.2369)

KAPPA INDEX OF AGREEMENT (KIA)

 Overall Kappa = 0.6869

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed
 vegetation, 6: Water, 7: Shadow

3. Error matrix for the result of classification with ECHO.

Error Matrix Analysis of EVALU (columns : truth) against ECH2501 (rows : mapped)

Class	1	2	3	4	5	6	7	Total	ErrorC
1	10622	2662	638	22	13	41	1202	15200	0.3012
2	5555	10265	4160	2056	138	182	1479	23835	0.5693
3	117	853	26886	1121	161	0	2	29140	0.0774
4	26	325	1079	1481	0	0	0	2911	0.4912
5	1	1	87	0	381	0	31	501	0.2395
6	26	25	53	0	0	1783	59	1946	0.0838
7	867	620	329	0	65	92	5890	7863	0.2509
Total	17214	14751	33232	4680	758	2098	8663	81396	
ErrorO	0.3829	0.3041	0.1910	0.6835	0.4974	0.1501	0.3201		0.2959

ErrorO = Errors of Omission (expressed as proportions)
 ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0026 (0.2933 - 0.2986)
 95% Confidence Interval = 0.0031 (0.2928 - 0.2991)
 99% Confidence Interval = 0.0041 (0.2918 - 0.3001)

KAPPA INDEX OF AGREEMENT (KIA)

 Overall Kappa = 0.6045

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed
 vegetation, 6: Water, 7: Shadow

4. Error matrix for the result of classification with divergence index.*

Error Matrix Analysis of EVALU (columns : truth) against DIV2501 (rows : mapped)

Class	1	2	3	4	5	6	7	Total	ErrorC
1	10464	4910	9916	688	28	121	4948	31075	0.6633
2	4719	6806	5997	265	277	35	937	19036	0.6425
3	230	397	8015	1312	16	0	32	10002	0.1987
4	91	311	1852	1897	22	0	0	4173	0.5454
5	3	53	193	1	53	3	13	319	0.8339
6	40	224	1319	19	309	1773	1439	5123	0.6539
7	163	81	2683	53	0	52	720	3752	0.8081
8	0	6	0	0	0	0	1	7	1.0000
Total	17214	14751	33232	4680	758	2098	8663	81396	
ErrorO	0.3339	0.4678	0.7326	0.5521	0.9248	0.1064	0.9110		0.5955

ErrorO = Errors of Omission (expressed as proportions)

ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0028 (0.6320 - 0.6375)

95% Confidence Interval = 0.0033 (0.6315 - 0.6381)

99% Confidence Interval = 0.0044 (0.6304 - 0.6391)

KAPPA INDEX OF AGREEMENT (KIA)

Overall Kappa = 0.2540

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed vegetation, 6: Water, 7: Shadow, 8: Unknown

5. Error matrix for the result of MHL classification with patch mean.

Error Matrix Analysis of EVALU (columns : truth) against MLH2501 (rows : mapped)

Class	1	2	3	4	5	6	7	Total	ErrorC
1	12766	2291	160	136	11	440	2672	18476	0.3090
2	2712	8936	467	14	88	18	367	12602	0.2909
3	190	959	30800	2019	353	8	25	34354	0.1035
4	19	365	561	2163	0	0	0	3108	0.3041
5	1	5	94	19	212	0	0	331	0.3595
6	0	10	40	0	0	1544	112	1706	0.0950
7	140	59	19	0	23	0	4807	5048	0.0477
Total	15828	12625	32141	4351	687	2010	7983	75625	
ErrorO	0.1935	0.2922	0.0417	0.5029	0.6914	0.2318	0.3978		0.1904

ErrorO = Errors of Omission (expressed as proportions)
 ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0025 (0.2453 - 0.2503)
 95% Confidence Interval = 0.0030 (0.2448 - 0.2507)
 99% Confidence Interval = 0.0039 (0.2439 - 0.2517)

KAPPA INDEX OF AGREEMENT (KIA)

 Overall Kappa = 0.7348

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed
 vegetation, 6: Water, 7: Shadow

6. Error matrix for the result of MHL classification with patch pdf.

Error Matrix Analysis of EVALU (columns : truth) against GAP2501 (rows : mapped)

Class	1	2	3	4	5	6	7	Total	ErrorC
1	13139	1827	234	104	17	157	1465	16943	0.2245
2	2193	9466	445	13	17	16	362	12512	0.2434
3	188	915	30708	1994	368	8	0	34181	0.1016
4	20	337	594	2285	0	0	0	3236	0.2939
5	3	9	82	2	275	0	25	396	0.3056
6	0	2	79	0	0	1843	0	1924	0.0421
7	285	25	20	0	15	0	6057	6402	0.0539
8	0	26	0	0	0	0	1	27	1.0000
Total	15828	12607	32162	4398	692	2024	7910	75621	
ErrorO	0.1699	0.2491	0.0452	0.4804	0.6026	0.0894	0.2343		0.1567

ErrorO = Errors of Omission (expressed as proportions)
 ErrorC = Errors of Commission (expressed as proportions)

90% Confidence Interval = 0.0024 (0.2141 - 0.2189)

95% Confidence Interval = 0.0028 (0.2137 - 0.2193)

99% Confidence Interval = 0.0037 (0.2128 - 0.2202)

KAPPA INDEX OF AGREEMENT (KIA)

Overall Kappa = 0.7828

Class values 1: Building, 2: Road, 3: Forest, 4: Lawn, 5: Shadowed
 vegetation, 6: Water, 7: Shadow, 8: Unknown

REFERENCES

- Albrecht, J., 1996. Geographic objects, and how to avoid them. In: Geographic Objects with Undetermined Boundaries (Burrough, P. A. and A. U. Frank, editors), Taylor & Francis Inc., Bristol, PA, pp. 325-331.
- Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer, 1976. *A land use and land cover classification system for use with remote sensor data*, Geological survey professional paper 964, United States Geological Survey, Washington, D.C.
- Anderson, T. W., 1984. An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, pp. 156-194.
- Atkinson, P. M., and P. J. Curran, 1997. Choosing an appropriate spatial resolution for remote sensing investigations. *Photogrammetric Engineering and Remote Sensing* 63 (12): 1345-1351.
- Barber, D. G., and E. F. LeDrew. 1991. SAR sea ice discrimination using texture statistics: a multivariate approach. *Photogrammetric Engineering and Remote Sensing* 57 (4): 385-395.
- Barzohar, M., M. Cohen, I. Ziskind, and D. B. Cooper, 1997. Fast robust tracking of curvy partially occluded roads in clutter in aerial images. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) (A. Gruen, E. P. Baltsavias, and O. Henricsson, editors), Birkhauser Verlag, Basel, pp. 277-286.
- Baumgartner, A., W. Eckstein, H. Mayer, C. Heipke, and H. Ebner, 1997. Context-supported road extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) (A. Gruen, E. P. Baltsavias, and O. Henricsson, editors), Birkhauser Verlag, Basel, pp. 299-308.
- Bezdek, J., 1993. Fuzzy models-what are they, and why? *IEEE Transactions on Fuzzy Systems* 1(1): 1-5.
- Bian L., and R. Butler, 1999. Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data. *Photogrammetric Engineering and Remote Sensing* 65 (1): 73-84.
- Bryant, J., 1990. AMOEBA clustering revisited. *Photogrammetric Engineering and Remote Sensing* 56 (1): 41-47.
- Cao, C., and N. S. Lam, 1997. Understanding the scale and resolution effects in remote sensing and GIS. In: Scale in Remote Sensing and GIS (D. A. Quattrochi, and M. F. Goodchild, editors), Lewis Publishers, New York, NY, pp. 57-72.

- Chiang, A. C., 1974. Fundamental Methods of Mathematical Economics. McGRAW-HILL BOOK COMPANY, New York, pp. 135-465.
- Cleynenbreugel, J. V., F. Fierens, and A. Oosterlinck, 1990. Delineating road structures on satellite imagery by a GIS-guided technique. *Photogrammetric Engineering and Remote Sensing* 56 (6): 893-898.
- Cohen, J. (1960), A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* XX:37-44.
- Coleman, G. B., and H. C. Andrews, 1979. Image segmentation by clustering. *Proceedings of the IEEE* 67 (5): 773-785.
- Collins, J. B. and C. E. Woodcock, 1999a. Modelling the distribution of cover fraction of a geophysical field. In: Advances in Remote Sensing and GIS Analysis (P. M. Atkinson and N. J. Tate), John Wiley and Sons, London, pp. 119-133.
- Collins, J. B., and C. E. Woodcock, 1999b. Geostatistical estimation of resolution dependent variance in remotely sensed images. *Photogrammetric Engineering and Remote Sensing* 65 (1): 41-50.
- Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37: 35-46.
- Curran P. J., 1988. The semivariogram in remote sensing: an introduction. *Remote Sensing of Environment* 24: 493-507.
- Cushnie J. L., 1987. The interactive effect of spatial resolution and degree of internal variability within land-cover types on classification accuracies. *International Journal of Remote Sensing* 8 (1): 15-29.
- Eastman, J. R., 1997. Idrisi for Windows User's Guide, version 2.0, Clark University, Worcester, MA.
- Elder, J. H., and S. W. Zucker, 1998. Local scale control for edge detection and blur estimation. *IEEE Transaction on Pattern Analysis and Intelligence* 20 (7): 699-716.
- ERDAS, 1999. ERDAS Field Guide, ERDAS, Atlanta, Georgia, 672 p.
- Ferro, C., 1998. Scale and Texture in Digital Image Classification, Unpublished MA thesis, West Virginia University, Morgantown WV, 52 p.
- Fjørtoft, R, A. Lopes, P. Marthon, and E. Cubero_Castan, 1998. An optimal multiedge detector for SAR image segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 36 (3): 793-802.

- Foody, G. M., 1999. The continuum of classification fuzziness in thematic mapping, *Photogrammetric Engineering and Remote Sensing*, 65, 443-451.
- Ford, S. J., and D. M. McKeown, Jr., 1992, Utilization of multispectral imagery for cartographic feature extraction. In Proceedings of the international forum on airborne multispectral scanning for forest and mapping (with emphasis on MEIS), Forestry Canada, Canada, pp. 129-143.
- Franklin, S. E., and B. A. Wilson, 1991. Spatial and spectral classification of remote sensing imagery. *Computers & Geosciences* 17(8): 1151-1172.
- Fung, T., and K. Chan, 1994. Spatial composition of spectral classes: a structural approach for image analysis of heterogeneous land-use and land-cover types. *Photogrammetric Engineering and Remote Sensing* 60 (2): 173-180.
- Gahegan, M., and J. Flack, 1996. A model to support the integration of image understanding techniques within a GIS. *Photogrammetric Engineering and Remote Sensing* 62 (5): 483-490.
- Gahegan, M., and J. Flack, 1999. The integration of scene understanding within a geographic information system: A prototype approach for agricultural applications. *Transactions in Geographic Information Systems* 3 (1): 31-49.
- Gong, P., D. J. Marceau, and P. J. Howarth, 1992. A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment* 40: 137-151.
- Gong, P., and P. J. Howarth, 1990. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering and Remote Sensing* 56 (1): 67-73.
- Gougeon, F. A., 1992. Individual tree identification from high resolution MEIS images. In: Proceedings of the international forum on airborne multispectral scanning for forest and mapping (with emphasis on MEIS), Forestry Canada, Canada, pp. 117-128.
- Gougeon, F. A., 1995a. Comparison of possible multispectral classification schemes for tree crowns individually delineated on high spatial resolution MEIS images. *Canadian Journal of Remote Sensing* 21(1): 1-9.
- Gougeon, F. A., 1995b. A crown following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing* 21(3): 274-284.
- Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3(6): 610-621.

- Haralick, R. M., and L. G. Shapiro, 1985. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing* 29: 100-132.
- Hay, G. J., K. O. Niemann, and D. G. Goodenough, 1997. Spatial thresholds, image-objects, and upscaling: a multiscale evaluation. *Remote Sensing of Environment* 62: 1-19.
- Hay, G. J., K. O. Niemann, and G. F. McLean. 1996. An object-specific image-texture analysis of H-resolution forest imagery. *Remote Sensing of Environment* 55: 108-122.
- Henebry, G. M., 1993. Detecting change in grasslands using measures of spatial dependence with Landsat TM data. *Remote Sensing of Environment* 46: 223-234.
- Hodgson, M. E., 1998. What size window for image classification? A cognitive perspective. *Photogrammetric Engineering and Remote Sensing* 64 (8): 797-807.
- Holyer, R. J., and S. H. Peckinpaugh, 1989. Edge detection applied to satellite imagery of the oceans. *IEEE Transactions on Geoscience and Remote Sensing* 27 (1): 46-56.
- Janssen, L. L. F., and M. Molenaar, 1995. Terrain objects, their dynamics and their monitoring by the integration of GIS and Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing* 33 (3): 749-758.
- Jensen, J. R., 1996. Introductory Digital Image Processing, Prince Hall, Upper Saddle River, New Jersey. York, NY, pp. 3-78.
- Jensen, J. R., and D. C. Cowen, 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing* 65 (5): 611-622
- Journal, A. G. and C. J. Hiujbregts, 1978. Mining Geostatistics, Academic Press, London, 600 p.
- Karniell, A., A. Melsels, L. Fisher, and Y. Arkin, 1996. Automatic extraction and evaluation of geological linear features from digital remote sensing data using a Hough transform. *Photogrammetric Engineering & Remote Sensing* 62 (5): 525-531.
- Kettig, R. L., and D. A. Landgrebe, 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics* GE-14(1): 19-26.
- Kloditz, C., A. van Boxtel, and E. Carfagna, 1998. Estimating the accuracy of coarse scale classification using high scale information. *Photogrammetric Engineering & Remote Sensing* 64 (2):127-133.

- Kramer, H. J. 1996. Observation of the Earth and its Environment: Survey of Missions and Sensors, Springer, Berlin, Germany, pp. 1-755.
- Kurosu, T., S. Uratsuka, H. Maeno, and T. Kozu, 1999. Texture statistics for classification of land use with multitemporal JERS-1 SAR single-look imagery. *IEEE Transactions on Geoscience and Remote Sensing* 37 (1): 227-235.
- Landgrebe D. and L. Biehl, 2000. An introduction to Multispec, version 1.2, Unpublished report, Purdue University, West Lafayette IN, 168 p. (<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Intro00.pdf>)
- Latty, R. S., R. Nelson, B. Markham, D. Williams, D. Toll, and J. Irons, 1985. Performance comparisons between information extraction techniques using variable spatial resolution data. *Photogrammetric Engineering and Remote Sensing* 51 (9): 1495-1470.
- Lea, S. M. and M. Lybanon, 1993. Automated boundary delineation in infrared ocean images. *IEEE Transactions on Geoscience and Remote Sensing* 31 (6): 1256-1260.
- Lillesand, T. M., and R. W. Kieffer, 1994. Remote sensing and image interpretation, John Wiley & Sons, Inc. New York, NY, pp. 169-178.
- McKeown, D. M., 1990. Toward automatic cartographic feature extraction. In: Mapping and Spatial Modelling for Navigation (Pau, L. F., editor), NATO ASI series F65: 149-180.
- Marceau, D. J., D. J. Gratton, R. A. Fournier, and J. Fortin, 1994a. Remote sensing and the measurement of geographical entities in a forested environment. 1. The scale and spatial aggregation problem. *Remote Sensing of Environment* 49: 93-104.
- Marceau, D. J., D. J. Gratton, R. A. Fournier, and J. Fortin, 1994b. Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment* 49: 105-117.
- Marr, D., 1982, Vision: A computational investigation into the human representation and processing of visual information, San Francisco, W. H. Freeman and Company, 397 p.
- Mason, D. C., D. G. Corr, A. Cross, D. C. Hogg, D. H. Lawrence, M. Petrou, and A. M. Taylor, 1988. The use of digital map data in the segmentation and classification of remotely-sensed images. *International Journal of Geographic Information Systems* 2 (3): 195-219.
- Matheron, G., 1963. Principles of geostatistics. *Economic Geology* 58: 1246-1266.

- Mcdevitt, R. J., and S. D. Peddada, 1998. An automated algorithm for cleaning and ordering the boundary points of a one-dimensional curve in a segmented image. *IEEE Transactions on Geoscience and Remote Sensing* 36 (1): 307-312.
- Meyer, P., K. Staenz, and K. I. Itten, 1996. Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography. *ISPRS Journal of Photogrammetry & Remote Sensing* 51: 5-16.
- Møller-Jensen, L. 1990. Knowledge-based classification of an urban area using texture and context information in Landsat-TM imagery. *Photogrammetric Engineering and Remote Sensing* 56(6): 899-904.
- Muerle, J. and D. Allen, 1968. Experimental evaluation of techniques for automatic segmentation of objects in a complex scene. In: Pictorial Pattern Recognition (G. Cheng, *et al.*, editors), Thompson, Washington D.C.
- Oliver, C. J., D. Blacknell, and R. G. White, 1996. Optimum edge detection in SAR. *IEE Proceedings of Radar and Sonar Navigation* 143 (1): 31-40.
- Oliver, C. J., and P. Lombardo, 1996. Simultaneous mean and texture edge detection in SAR clutter. *IEE Proceedings of Radar and Sonar Navigation* 143 (6): 391-399.
- Pax-Lenney, M., and C. E. Woodcock, 1997. The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote sensing of Environment* 61: 210-220.
- Pouch, G. W., and D. J. Campagna, 1990. Hyperspherical direction cosine transformation for separation of spectral and illumination information in digital scanner data. *Photogrammetric Engineering and Remote Sensing* 56 (4): 475-479.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1999. Numerical Recipes in Fortran 77. The Art of Scientific Computing, New York, Cambridge University Press, pp. 205-649.
- Price, J. C., 1994. How unique are spectral signatures? *Remote Sensing of Environment* 49: 181-186.
- Richards, J. A., 1994. Remote Sensing Digital Image Analysis An Introduction, Berlin, Springer-Verlag, pp. 181-291.
- Roberts R. J., 1965. Machine perception of three dimensional solids. In: Optical and electro-optical information processing (J. T. Tippett, D. A. Berkowitz, L. C. Clapp, C. J. Koester, and A. Vanderburgh, Jr., editors), MIT Press, Cambridge, MA, pp. 159-197.

- Roberts, S. A., M. N. Gahegan, J. Hogg, and B. Hoyle, 1991. Application of object-oriented databases to geographic information systems. *Information and Software technology* 33 (1): 38-46.
- Roberts, S. A., and M. N. Gahegan, 1993. An object-oriented geographic information systems shell. *Information and Software technology* 35 (10): 561-572.
- Ryherd, S., and C. Woodcock. 1996. Combining spectral and texture data in the segmentation of remotely sensed images. *Photogrammetric Engineering and Remote Sensing* 62(2): 181-194.
- Sahar, L., and A. Krupnik, 1999. Semiautomatic extraction of building outlines from large-scale aerial images. *Photogrammetric Engineering and Remote Sensing* 65(4): 459-465.
- Shufelt, J. A., 1999. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (4): 311- 325.
- St-Onge, B. A., and F. Cavayas. 1997. Automated forest structure mapping from high resolution imagery based on directional semivariogram estimates. *Remote Sensing of Environment* 61: 82-95.
- Stow, D., A. Hope, A. T. Nguyen, S. Phinn, and C. A. Benkelman, 1996. Monitoring detailed land surface changes using an airborne multispectral digital camera system. *IEEE Transactions on Geoscience and Remote Sensing* 34 (5): 1191-1203.
- Strahler, A. H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment* 10: 135-163.
- Strahler, A. H., C. E. Woodcock, and J. A. Smith, 1986. On the nature of models in remote sensing. *Remote Sensing of Environment* 20: 121-139.
- Swain, P. and S. Davis, 1978. Remote Sensing: The Quantitative Approach, McGraw Hill, N.Y. 369 p.
- Tang, A. Y., T. M. Adams, and E. L. Usery, 1996. A spatial data model design for feature-based geographical information systems. *International Journal of Geographic Information Systems* 10 (5): 643-659.
- Teillet, P. M., K. Staenz, and D. J. Williams, 1997. Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sensing of Environment* 61: 139-149.
- Ton, J., J. Stcklen, and A. K. Jain, 1991. Knowledge-based segmentation of Landsat images. *IEEE Transactions on Geoscience and Remote Sensing* 29 (2): 222-231.

- Tou, J. T. and R. C. Gonzalez, 1974. Pattern Recognition Principles, Addison-Wesley Publishing Co, Reading, MA.
- Touzi, R., A. Lopes, and P. Bousquet, 1988. A statistical and geometrical edge detector for SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 26 (6): 764-773.
- Treitz, P. and P. Howarth, 2000. Integrating spectral, spatial and terrain variables for forest ecosystem classification. *Photogrammetric Engineering and Remote Sensing* 66(3): 305-317.
- Trinder, J. C., Y. Wang, A. Sowmya, and M. Palhang, 1997. Artificial intelligence in 3-D feature extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) (A. Gruen, E. P. Baltsavias, and O. Henricsson, editors), Birkhauser Verlag, Basel, pp. 257-266.
- Tucker, C. J., B. N. Holben, J. H. Elgin and J. E. McMurtrey, 1981. Remote sensing of total dry matter accumulation in winter wheat. *Remote Sensing of Environment* 13, 461.
- Usery, E. L. 1993. Category theory and the structure of features in geographic information systems. *Cartography and Geographic Information Systems* 20 (1): 5-12.
- Usery, E. L. 1996. A feature-based geographic information system model. *Photogrammetric Engineering and Remote Sensing* 62(7): 833-838.
- USGS, 1997. American National Standard for Information System – Spatial Data Transfer Standard (SDTS) – Part 2, Spatial Feature, Unpublished report. American National Standards Institute, Inc. (http://mcmcweb.er.usgs.gov/sdts/SDTS_standard_nov97/part1bo8.html)
- Vincent, L., and P. Soille, 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (6): 583-598.
- Wang, L., and D. C. He, 1990. A new statistical approach for texture analysis. *Photogrammetric Engineering and Remote Sensing* 56: 61-66.
- Warner, T. A., J. Y. Lee, and J. B. McGraw, 1999. Delineation and identification of individual trees in the Eastern Deciduous Forest. In: International Forum on Automated Interpretation of High Spatial Resolution Imagery for Forestry (D. Hill and D. Leckie, editors), February 10-12, 1998, Victoria B.C. pp. 81-91.

- Warner, T. A., and M. C. Shank, 1997a. An evaluation of the potential for fuzzy classification of multispectral data using artificial neural networks. *Photogrammetric Engineering and Remote Sensing* 63 (11): 1285-1294.
- Warner, T. A., and M. C. Shank, 1997b. Spatial autocorrelation analysis of hyperspectral imagery for feature selection. *Remote Sensing of Environment* 60:58-70.
- Welch, R., 1982. Spatial resolution requirements for urban studies. *International Journal of Remote Sensing* 3 (2): 139-146
- Wharton, S. W., 1987. A spectral-knowledge-based approach for urban land-cover discrimination, *IEEE Transactions on Geoscience and Remote Sensing* GE-25. No. 3: 272-282.
- Woodcock, C. E., and V. J. Harward, 1992. Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing of Environment* 13 (16): 3167-3187.
- Woodcock, C. E., and A. H. Strahler. 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21: 311-332.
- Woodcock, C. E., A. H. Strahler and D. L. B. Jupp. 1988a. The use of variograms in remote sensing, I: scene models and simulated images. *Remote Sensing of Environment* 25: 323-348.
- Woodcock, C. E., A. H. Strahler and D. L. B. Jupp. 1988b. The use of variograms in remote sensing, II: real digital images. *Remote Sensing of Environment* 25: 349-379.
- Worboys, M. F., 1994. Object-oriented approaches to geo-referenced information, *International Journal of Geographic Information Systems* 8 (4): 385-399.