

DEVELOPMENT OF URBAN AREA GEOSPATIAL INFORMATION  
PRODUCTS FROM HIGH RESOLUTION SATELLITE IMAGERY  
USING ADVANCED IMAGE ANALYSIS TECHNIQUES

---

A Dissertation  
presented to  
the Faculty of the Graduate School  
University of Missouri-Columbia

---

In Partial Fulfillment  
Of the Requirements for the Degree

Doctor of Philosophy

---

by  
AARON K. SHACKELFORD

Dr. Curt H. Davis, Dissertation Supervisor

DECEMBER 2004

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled.

DEVELOPMENT OF URBAN AREA GEOSPATIAL INFORMATION  
PRODUCTS FROM HIGH RESOLUTION SATELLITE IMAGERY  
USING ADVANCED IMAGE ANALYSIS TECHNIQUES

Presented by Aaron K. Shackelford

A candidate for the degree of Doctor of Philosophy of Electrical Engineering

And herby certify that in their opinion it is worthy of acceptance.

Dr. Curt H. Davis



---

Dr. Craig A. Kluever



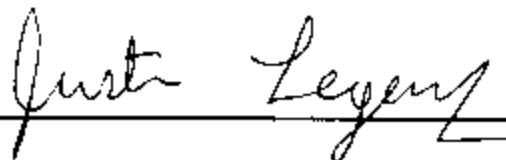
---

Dr. Dominic K. C. Ho



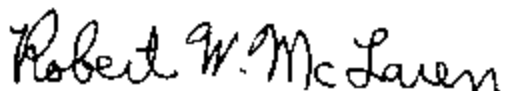
---

Dr. Justin Legarsky



---

Dr. Robert W. McLaren



---

*For my wife*

## ACKNOWLEDGMENTS

Throughout my many years as a student at the University of Missouri-Columbia, many individuals have been instrumental in getting me to the point where I now find myself. It would be impossible to thank each person by name, so I will just mention a few here. First and foremost, I wish to thank my advisor, Dr. Curt Davis. His advice and guidance were invaluable during my time as a doctoral student. It was truly a pleasure to work with him on this research project. I also wish to thank the members of my doctoral committee, Dr. Craig Kluever, Dr. Dominic Ho, Dr. Justin Legarsky, and Dr. Robert McLaren, for their valuable advice.

This research was funded by the NASA Graduate Student Researchers Program. I would like to thank my NASA technical advisor, Dr. Bruce Davis at the Stennis Space Center, for his support of this project.

Last, but certainly not least, I would like to thank my family. First, I wish to thank my wife, Lisa, for her love and for the many sacrifices she made so that I could pursue my graduate studies. Without her constant support and encouragement, it would have been impossible to complete this dissertation. Finally, I would like to thank my parents, Keith and Deborah, and my brother, Grant, for their unceasing love and support.



# TABLE OF CONTENTS

<b>Acknowledgements</b> .....	ii
<b>List of Illustrations</b> .....	vi
<b>List of Tables</b> .....	ix
<b>Abstract</b> .....	xii
<b>1. Introduction</b>	
Geospatial Information Products and Applications .....	1
Satellite Imagery .....	5
Satellite Image Analysis and Classification .....	20
Summary of Related Work .....	31
Summary and Research Objectives .....	38
<b>2. Fuzzy Pixel-Based Urban Land Cover Classification</b>	
Introduction .....	43
Maximum Likelihood Classification Study .....	44
Spatial Feature Extraction .....	51
Hierarchical Fuzzy Classification .....	57
Fuzzy Pixel Classifier Results .....	63
Summary .....	70
<b>3. Combined Pixel and Object-Based Fuzzy Urban Land Cover Classification</b>	
Introduction .....	72
Segmentation .....	74
Fuzzy Logic Object Classification .....	77

Fuzzy Logic Object Classifier Results .....	89
Summary .....	98
<b>4. Automated Road Network Extraction</b>	
Introduction .....	101
Accuracy Assessment .....	102
Road Extraction from Land Cover Classification .....	104
Road Extraction from High Resolution Imagery .....	108
Summary .....	122
<b>5. Automated 2D Building Footprint Extraction</b>	
Introduction .....	126
Differential Morphological Profile .....	127
Building Extraction .....	133
Test Results .....	138
Summary .....	142
<b>6. Fully- Automated Urban Land Cover Classification</b>	
Introduction .....	144
Fully Automated Training Data Generation .....	145
Self-Supervised Urban Land Cover Classification .....	150
Test Results .....	155
Summary .....	160
<b>7. Summary and Future Work</b>	
Research Summary .....	170
Future Work .....	176

<b>Reference List</b> .....	180
<b>VITA</b> .....	186

## LIST OF ILLUSTRATIONS

<b>Figure</b>	<b>Page</b>
1-1. Panchromatic USGS DOQQ .....	6
1-2. Color NIMA digital orthophoto .....	7
1-3. False color Landsat TM image .....	10
1-4. False color IKONOS MS image .....	11
1-5. IKONOS PAN sharpening example .....	15
1-6. Landsat ETM+ PAN of commercial area .....	17
1-7. DOQQ orthophoto of commercial area .....	17
1-8. NIMA orthophoto of commercial area .....	18
1-9. IKONOS PAN, MS, and PS-MS of commercial area .....	18, 19
2-1. IKONOS PAN of Columbia, Missouri .....	46
2-2. Maximum likelihood classification of suburban and urban areas .....	49
2-3. Texture classification example .....	52
2-4. 2-D spatial contextual feature extraction .....	55
2-5. Length-width contextual measures of suburban and urban areas .....	57
2-6. Block diagram of hierarchical fuzzy classification scheme .....	59
2-7. Crisp output of fuzzy classifier for Columbia, Missouri imagery .....	65
2-8. Crisp output of fuzzy classifier for suburban and urban areas .....	66
3-1. 1 m resolution PS-MS IKONOS image of dense urban area .....	77

<b>Figure</b>	<b>Page</b>
3-2. Segmentation of dense urban image .....	77
3-3. Morphological skeleton shape processing example .....	81
3-4. 1 m PS-MS IKONOS urban test site .....	90
3-5. Object-based classification of dense urban area .....	92
3-6. Fuzzy logic object-based classification of urban test site .....	95
3-7. Hierarchical pixel-based fuzzy classification of urban test site .....	96
3-8. Fuzzy logic object-based classification of urban test site with suburban building discrimination .....	97
3-9. 1 m PS-MS IKONOS suburban test site .....	98
3-10. Fuzzy logic object-based classification of suburban test site .....	98
4-1. IKONOS PS-MS images of suburban and urban subsets .....	105
4-2. Road network extracted from land cover classification by skeletonization .....	106
4-3. Road network extracted from land cover classification by iterative algorithm .....	109
4-4. Fully automated road network extraction algorithm flow chart .....	113
4-5. Road segment growth example .....	117, 118
4-6. Road network extracted from IKONOS by fully automated algorithm .....	119
4-7. Road network extracted from NIMA orthophoto commercial test site by fully automated algorithm .....	122

<b>Figure</b>	<b>Page</b>
4-8. Road network extracted from NIMA orthophoto suburban test site by fully automated algorithm .....	123
5-1. Opening differential morphological profile .....	131
5-2. Closing differential morphological profile .....	132
5-3. Building extraction algorithm flow chart .....	135
5-4. Indirect building extraction via shadow identification example .....	139
5-5. Automated building extraction results for dense urban area .....	140
6-1. Automatically generated training data for <i>Road</i> and <i>Building</i> classes .....	154
6-2. Automatically generated training data for <i>Grass</i> and <i>Tree</i> classes .....	155
6-3. Self-supervised classification flow chart .....	157
6-4. Self-supervised fuzzy pixel-based classification of dense urban area subset .....	163
6-5. Self-supervised fuzzy object-based classification of dense urban area subset .....	163
6-6. Self-supervised fuzzy pixel-based classification of urban test site .....	165
6-7. Self-supervised fuzzy object-based classification of urban test site .....	166
6-8. Self-supervised fuzzy object-based classification of urban test site with suburban building discrimination .....	167
6-9. Self-supervised fuzzy object-based classification of suburban test site .....	168
7-1. Urban area geospatial information products extracted by fully automated techniques .....	171

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
1-1. Landsat TM and ETM+ spectral bands .....	9
1-2. IKONOS spectral bands .....	11
1-3. QuickBird spectral bands .....	12
2-1. Maximum likelihood classification accuracy .....	48
2-2. Confusion matrix for maximum likelihood classification of PS-MS of Columbia, MO .....	50
2-3. Confusion matrix for maximum likelihood classification of PS-MS of Springfield, MO .....	50
2-4. Maximum likelihood classification accuracies with added texture information .....	54
2-5. Fuzzy classifier overall accuracies of PS-MS .....	64
2-6. Confusion matrix for fuzzy classification of PS-MS of Columbia, MO .....	66
2-7. Confusion matrix for fuzzy classification of PS-MS of Springfield, MO .....	67
2-8. Fuzzy classifier accuracies of 4 m MS of Columbia, MO .....	68
2-9. Confusion matrix for maximum likelihood classification of 4 m MS of Columbia, MO .....	68
2-10. Confusion matrix for fuzzy classification of 4 m MS of Columbia, MO .....	69
3-1. Classification accuracies for different combinations of object-based features .....	91
3-2. Confusion matrix for object-based classification of urban area test site .....	93

<b>Table</b>	<b>Page</b>
3-3. Confusion matrix for pixel-based classification of urban area test site .....	94
4-1. Road extraction performance measures from skeletonization and pruning .....	106
4-2. Road extraction performance measures from iterative algorithm .....	108
4-3. Road extraction performance measures from fully automated extraction of IKONOS test sites .....	120
4-4. Fully automated algorithm parameters for IKONOS test sites .....	120
4-5. Road extraction performance measures from fully automated extraction of orthophoto test sites .....	121
4-6. Fully automated algorithm parameters for orthophoto test sites .....	122
4-7. Comparison of results reported in automated road extraction literatures .....	125
5-1. Object level building extraction performance measures .....	142
5-2. Pixel level building extraction performance measures .....	142
6-1. Accuracies of maximum likelihood classifications .....	158
6-2. Accuracies of fuzzy pixel-based classifications .....	158
6-3. Accuracies of fuzzy object-based classification .....	159
6-4. Confusion matrix for self-supervised maximum likelihood classification .....	160
6-5. Confusion matrix for self-supervised fuzzy pixel-based classification .....	160
6-6. Confusion matrix for self-supervised fuzzy Object-based classification .....	161



<b>Table</b>		<b>Page</b>
6-7.	Confusion matrix for supervised maximum likelihood classification .....	161
6-8.	Confusion matrix for supervised fuzzy pixel-based classification .....	161
6-9.	Confusion matrix for supervised fuzzy object-based classification .....	162

DEVELOPMENT OF URBAN AREA GEOSPATIAL INFORMATION  
PRODUCTS FROM HIGH RESOLUTION SATELLITE IMAGERY  
USING ADVANCED IMAGE ANALYSIS TECHNIQUES

Aaron K. Shackelford

Dr. Curt H. Davis, Dissertation Supervisor

ABSTRACT

There are numerous applications for urban area geospatial information products at both the local and federal government levels, as well as in the defense and intelligence communities. These applications include urban area growth assessment, emergency response and management, national security, military combat support, and homeland security. The latest generation of commercial satellite imaging sensors have a number of characteristics (e.g. high spatial resolution, multispectral bands, and quick revisit time) that make them ideal data sources for urban area applications. The goal of this doctoral research was to develop advanced automated and semi-automated image analysis and classification techniques for the extraction of urban area geospatial information products from commercial high-resolution satellite imagery. We developed two semi-automated fuzzy urban land cover classification approaches, as well as fully automated techniques for road network and 2-D building footprint extraction. The semi-automated classifiers are able to classify urban land cover significantly more accurately than traditional classification approaches, such as maximum likelihood. By utilizing fully automated feature extraction techniques for training data generation, a self-supervised urban land

cover classification approach was also developed. The self-supervised classifier is significantly more accurate than traditional classification approaches, and unlike traditional approaches it is fully automated. The development of automated and semi-automated techniques for generation of urban geospatial information products is of high importance not only for the many applications for which they can be used but also because the large volume of data collected by satellite sensors exceeds the human capacity of trained image specialists to analyze. In addition, many applications, especially those for the military and intelligence communities, require near real time exploitation of image data. Both the quantity of data and the speed with which that data must be analyzed illustrate the need for automated geospatial information extraction techniques.

# **Chapter 1**

## **Introduction**

### **1.1 Geospatial Information Products and Applications**

Urban growth and economic development place heavy demands on local governments to seek better planning and management approaches to deal with the numerous socioeconomic and environmental problems associated with urbanization. Natural resources (greenspace, wooded areas, agriculture lands, etc.) and existing infrastructure (transportation networks, sewage and water utilities, parks, etc.) are stressed as they are utilized beyond the capacity for which they were originally designed. Decision makers within various local government bodies must deal with a wide range of issues that routinely have economic, social, and political consequences. In addition, state and federal governments have issued an increasing number of regulations that mandate the monitoring and tracking of numerous items by local governments. These factors are occurring at such a rate that city and county resources, both labor and fiscal, cannot keep pace. Timely, accurate, and up-to-date geospatial information products are required by elected officials in these local governments to support policy decisions on issues that are often interrelated and can span several political boundaries. Growth assessment, infrastructure inventory and planning, environmental assessment, and risk management impact and drive policy decisions for these managers. Unfortunately, their decision-

making processes are often forced to rely on inaccurate data and/or outdated baseline information.

Many local governments are investing significant financial resources to implement Geographic Information Systems (GIS) to assist in planning and management activities to effectively deal with the multitude of issues caused by increasing urbanization. GIS can be used for a variety of tasks including map production, visualization to support geographic problem solving, spatial data analysis and modeling, engineering studies, and inventory management. At its core, GIS is an integrated database where each piece of information has a geographic location associated with it. The power of GIS is that the information is displayed in a spatial format and can be queried by spatial location. GIS data are organized by layers, which are collections of geographic entities of the same geometric type (points, lines, polygons, and 2-D raster imagery). For a detailed description of GIS and its many applications see Longley *et al.* [2001]. Typical layers found in a local government GIS systems include: digital image basemap, land use zoning, political boundaries, parcel maps, land cover, road network, building footprints, utility networks (water, sewage, electricity, etc.), topography, and green space. The digital image basemap is normally acquired through aerial photography or more recently satellite remote sensing. The image basemap is a very important layer in a local government GIS as several other information layers such as the land cover, road network, building footprints, and green space layers can be extracted from it [Jensen and Cowen, 1999]. The most widely used method for geospatial information extraction from imagery is manual digitization, which is both time consuming and expensive. Therefore,

the development of techniques for automated and/or semi-automated extraction of urban geospatial information products from remote sensing imagery is of great importance.

At the national level, the United States Geological Survey (USGS) is developing The National Map, which is “a database of continuously maintained base geographic information for the United States and its territories that will serve as the nation’s primary map for the 21<sup>st</sup> century” [USGS, 2001]. The National Map will provide accurate, current, and nationally consistent digital geospatial data and topographic maps. It will contain high-resolution orthorectified digital imagery, high-resolution surface elevation data, several vector feature data layers, geographic names for physical and cultural features, and land cover classification maps. The vector feature data includes layers for hydrography, transportation networks, building footprints, and boundaries for public and private lands. Potential uses for the image and feature data include urban growth planning, land and natural resource management, environmental protection, emergency response and management, and homeland security applications.

The USGS vision for The National Map includes near real time adaptation of the map to changes rather than cyclical inspection and revision. The ultimate goal is that changes will be recorded in the map within 7 days of an event. When a map update is needed, many of the feature vector layers will require new digital imagery of the area to be acquired and new features to be extracted. Manual digitization of features will not be sufficient to realize this near real time adaptation of The National Map to landscape changes, highlighting the need for automated extraction of these features.

In addition to civilian government applications, there are many defense and intelligence applications for geospatial information products. The National Imagery and

Mapping Agency (NIMA) was formed in 1996 to combine the then separate geospatial mapping (DMA) and imagery (NPIC) communities into a single agency focused on geospatial intelligence. NIMA defines geospatial intelligence as “the exploitation and analysis of imagery and geospatial information to describe, assess, and visually depict physical features and geographically referenced activities on the Earth” [NIMA, 2003]. The convergence of these two communities was due to the shifting national threat environment, the evolving nature of conflict, and advancement of technology. The applications of geospatial intelligence are informing statecraft, supporting military combat operations, homeland security and civil support, and intelligence collaboration. The discipline of geospatial intelligence utilizes a wide variety of remote sensing technologies such as optical imagery (panchromatic, multispectral, hyperspectral, etc.), synthetic aperture radar (SAR) and interferometric SAR (InSAR) data, and laser imaging systems on both airborne and satellite-based platforms. Geospatial information derived from these sensors are an integral source of information for geospatial intelligence analysis.

Following the tragic events of September 11, 2001, NIMA accelerated the convergence of various analytic tradecrafts (cartography, geospatial analysis, imagery analysis, geodesy, etc.) into a new intelligence discipline known as geospatial intelligence or GeoINT. With the signing of the FY 2004 Defense Authorization Bill, NIMA officially became the National Geospatial-Intelligence Agency (NGA) on November 24, 2003. The new name, NGA, is the latest step in an ongoing transformation of the agency to better serve the geospatial intelligence needs of the federal government to support homeland and national security. The expanded use of GeoINT in battlefield preparation,

precision targeting, and combat support during Operation Iraqi Freedom shortened the duration of the war and saved military as well as civilian lives.

## **1.2 Satellite Imagery**

There are a variety of commercially available image sources that can be used for urban applications. A brief review of these different sources is presented here. Imagery of urban areas must have a spatial resolution in the meter to sub-meter range to be useful for many of the different types of applications described in the previous section [Jenson, 1999]. The spatial resolution of commercially available satellite imagery did not increase to the level of detail required for urban analysis until recent years, so orthorectified digital images derived from aerial photography have traditionally been used for image basemaps in local government GIS systems. Therefore, in addition to satellite-based sensors, several aerial photography data sources will also be reviewed.

### *A. Aerial Photography*

Digital Orthophoto Quarter Quadrangles (DOQQ) produced by the USGS are the most widely available high-resolution digital orthoimage dataset available to local governments in the U.S. A digital orthophoto is a digital image generated from an aerial photograph or remotely sensed image in which the distortions due to terrain and sensor displacements have been removed giving it the positional accuracy and geometric qualities of a map [USGS, 1996]. Map-quality positional accuracy is crucial for imagery in a GIS because the basemap is of little value if the vector layers (parcel boundaries, road centerlines, utility lines, etc.) do not line up with the corresponding features in the





**Figure 1-1.** Panchromatic USGS DOQQ of St. Charles, Missouri (1 m resolution).

imagery. DOQQs are available as either panchromatic, natural color, or color infrared images with 1-m ground resolution and are derived primarily from National Aerial Photography Program (NAPP) imagery. The NAPP began in 1987 with the mission to acquire and archive photographic coverage of the United States at 1:40,000 scale with either color infrared or panchromatic aerial photography [Light, 1993]. The NAPP imagery is acquired on an approximately five-year cycle. DOQQs have an 8-bit per pixel dynamic range, and an example of a panchromatic DOQQ of St. Charles, Missouri is shown in Fig. 1-1. The DOQQs are used as the base image source for the USGS National Map over the majority of the U.S.



**Figure 1-2.** True color NIMA digital orthophoto of St. Charles (0.3 m (1 ft) resolution).

Another source of digital orthophotos available to local governments comes from a joint program between NIMA/NGA and the USGS to acquire high-resolution digital imagery for 133 major urban cities in the U.S. The USGS is utilizing the orthoimages as the base image layer for their National Map program [USGS, 2001] in these selected urban areas. NIMA/NGA is also using the imagery as a base image layer in conjunction with classified critical infrastructure data layers for homeland security applications. The

digital orthophotos are produced at 0.3 m (1 ft) resolution from natural color photography, resulting in a three band digital image (red, green, blue) with 8-bit per pixel quantization. An example of this imagery from St. Charles is shown in Fig. 1-2.

Both of the digital orthophoto types mentioned above are produced through precision digital scanning of analog aerial film photography. The production of multispectral (MS) data from natural-color or color infrared photography is done by filtering the photograph during the scanning process to capture only the portion of the spectrum in the band of interest. This process is repeated for each spectral band as desired. Because of losses inherent in the photographic film, filtering, and scanning process, MS digital orthophotos produced in this manner are of a lower quality than data from a MS digital imaging sensor, where separate data is collected for each spectral band. As digital aerial imaging sensors are a relatively new technology, they are not yet in widespread use.

### *B. Landsat*

The first widely available source of satellite-based remotely sensed imagery was the Landsat Program, which began in 1972 with the launch of the Landsat 1 satellite [USGS, 2003]. The program currently has two operational satellites, Landsat 5, launched in 1984, and Landsat 7, launched in 1999. Landsat 5 carries the multispectral scanning (MSS) sensor and thematic mapper (TM) sensors, and Landsat 7 carries the enhanced thematic mapper plus (ETM+) sensor. The TM sensor and the ETM+ sensor both collect data in seven spectral bands, as summarized in Table 1-1 below.

**Table 1-1**  
**Spectral bands for Landsat TM and ETM+ sensors**

<b>Band #</b>	<b>Spectral Band</b>	<b>Wavelength (μm)</b>
1	blue-green	0.45 - 0.52
2	green	0.52 - 0.60
3	red	0.63 - 0.69
4	near infrared	0.76 - 0.90
5	mid infrared	1.55 - 1.75
6	thermal infrared	10.40 – 12.50
7	far infrared	2.08 – 2.35
8*	PAN	0.52 – 0.90

\* PAN band only on ETM+ sensor.

Bands 1-5 and 7 have a spatial resolution of 30 m. The TM thermal infrared band (band 6) has a spatial resolution of 120 m and the ETM+ band 6 has a spatial resolution of 60 m [USGS, 1999]. The panchromatic (PAN) band present on the ETM+ sensor has a 15-m spatial resolution and collects energy across bands 2-5, making it ideal for image sharpening techniques. Landsat 5 and 7 both have 16-day orbit cycles, however their orbits are offset allowing for an 8-day repeat coverage from one of the satellites.

TM and ETM+ data have many applications in the areas of global change research, agriculture, forestry, geology, resource management, geography, mapping, water quality, and oceanography [USGS, 2003]. However, their use in urban applications is very limited due to the medium-scale spatial resolution (15-30 m). At this resolution, the presence of man-made structures or built-up areas can be detected, but mapping of detailed features such as individual roads and buildings in the urban environment is not possible. A 30-m resolution TM image of St. Charles, MO is shown in Fig. 1-3.



**Figure 1-3.** False color 7,5,3 band combination Landsat TM image of St. Charles (30 m resolution).

### *C. Commercial High Resolution Satellites*

High-resolution satellite imagery was not commercially available until 2000 when Space Imaging's IKONOS satellite became operational after its launch in late 1999. IKONOS data contain four MS bands with a 4 m spatial resolution and a PAN band with 1 m resolution as summarized in the Table 1-2 below [Space Imaging, 2003].



**Table 1-2**  
**Spectral bands of IKONOS commercial remote sensing satellite**

<b>Spectral Band</b>	<b>Wavelength (<math>\mu\text{m}</math>)</b>	<b>Spatial Resolution (m)</b>
blue	0.45 - 0.52	4.0
green	0.51 - 0.60	4.0
red	0.63 - 0.70	4.0
near infrared	0.76 - 0.85	4.0
PAN	0.45 - 0.90	1.0



**Figure 1-4.** False color infrared IKONOS MS image of Columbia, Missouri (4 m resolution).

The PAN band collects data across all four of the MS channels, making it ideal for pan-sharpening techniques. The IKONOS satellite allows up to  $30^{\circ}$  off-nadir viewing, resulting in quick revisit times of approximately 3 days, depending on latitude. The data

is stored with an 11-bit per pixel dynamic range allowing increased discrimination of subtle spectral differences between objects. The 1-m spatial resolution of the PAN band is the same as that of the DOQQs, however the 11-bit dynamic range provides superior image contrast and quality. The high-resolution, quick revisit times and large dynamic range of IKONOS data make it ideal for urban applications. In addition, it has been shown that this imagery is a cost effective alternative to aerial photography for the generation of digital image basemaps for use by local governments [Davis and Wang, 2003]. An IKONOS image of Columbia, Missouri is shown in Fig. 1-4. IKONOS is the primary data source utilized in the analyses presented in the following sections of this dissertation.

Data acquired by the QuickBird satellite currently provide the highest spatial resolution satellite imagery commercially available. DigitalGlobe, Inc. launched the QuickBird satellite in 2001 and it became operational in early 2002. Like IKONOS, QuickBird consists of four MS bands and a PAN band, as summarized in Table 1-3 [DigitalGlobe, 2003].

**Table 1-3**  
**Spectral bands of QuickBird commercial remote sensing satellite**

<b>Spectral Band</b>	<b>Wavelength (<math>\mu\text{m}</math>)</b>	<b>Spatial Resolution (m) (nadir)</b>	<b>Spatial Resolution (m) (25<sup>0</sup> off-nadir)</b>
blue	0.45 - 0.52	2.44	2.88
green	0.52 - 0.60	2.44	2.88
red	0.63 - 0.69	2.44	2.88
near infrared	0.76 - 0.90	2.44	2.88
PAN	0.45 – 0.90	0.61	0.72

As seen in Table 1-3, the spatial resolution of imagery available from QuickBird varies according to the satellite viewing angle. As was the case with IKONOS data, the PAN

band captures energy over the entire spectrum contained in the four MS bands, making it ideal for use in pan sharpening applications. QuickBird allows for up to  $30^{\circ}$  off-nadir viewing, resulting in quick revisit times of at most 3.5 days, depending on latitude. The data is stored with an 11-bit per pixel dynamic range. Like data from the IKONOS satellite, the bit-depth, resolution, revisit times and spectral coverage make QuickBird imagery ideal for urban applications.

In addition to operating the QuickBird satellite, DigitalGlobe was awarded the \$500 million NextView contract by NIMA/NGA on September 30, 2003. The NextView contract will enable the development of a next-generation commercial, high-resolution imaging satellite and represents a new level of partnering between U.S. government agencies and the U.S. commercial remote sensing industry [Barnes, 2003]. Through this contract, NIMA/NGA and its military customers will have greater access to and priority acquisition of imagery, in addition to participation in the development cycle of the next-generation sensors. Although the exact specifications of DigitalGlobe's new satellite sensor, WorldView, have not yet been announced, the spatial resolution will be between 0.25 - 0.5 m and it will have eight MS bands, reaching into the low end of the infrared spectrum. WorldView will have the ability to acquire two to four times the daily amount of imagery compared to the QuickBird satellite. This is accomplished by flying the sensor at a higher altitude and increased onboard storage capacity. In addition, the system will be able to deliver an image to customers in less than 90 minutes from the time of acquisition. The NextView contract demonstrates the commitment of both NIMA/NGA and the intelligence community to the operational use of commercial satellite imagery for GeoINT applications.



#### *D. Image Pan Sharpening*

In the discussion of the Landsat, IKONOS, and QuickBird satellites, it was mentioned that the PAN bands of these sensors can be used for pan-sharpening applications. Pan-sharpening (also referred to as data fusion) is the process of increasing the spatial resolution of lower resolution MS bands using the higher resolution PAN band. This produces imagery with the spatial resolution of the PAN band and the spectral characteristics of the MS bands. The fact that the spectrum of the PAN band covers the range of the MS bands is crucial for the success of pan-sharpening techniques, because the PAN band must contain information from each of the MS bands if it is to accurately sharpen them. There are a variety of pan-sharpening algorithms, including wavelet-based techniques [Ranchin and Wald, 2000], principle components analysis, and arithmetic methods such as color normalization [Vrabel, 1996]. Wavelet-based techniques extract the high-frequency content from the PAN band with a wavelet transform and inject this information into the MS bands. Pan sharpening based on principle components analysis is accomplished by computing a transform to uncorrelate the MS bands, resulting in an uncorrelated dataset with the same number of bands as the MS data [Jenson, 1996]. The first principle component, representing brightness, is then replaced with the PAN band and a reverse principle components transform is applied. The color normalization technique multiplies each of the MS bands by the higher resolution PAN band, and these values are normalized by the sum of the MS bands. The MS bands must be resampled to the resolution of the PAN band before any of these techniques can be applied. The color normalization transformation from Vrabel [1996] is calculated as

$$PSMS_i = \frac{(MS_i + 1.0) * (PAN + 1.0) * 3.0}{\sum_{j=1}^b MS_j + 3.0} \quad (1-1)$$

where  $PSMS_i$  is band  $i$  of the pan-sharpened MS (PS-MS) imagery,  $MS_i$  is band  $i$  of the lower resolution MS imagery, and  $b$  is the number of MS bands in the dataset. Because



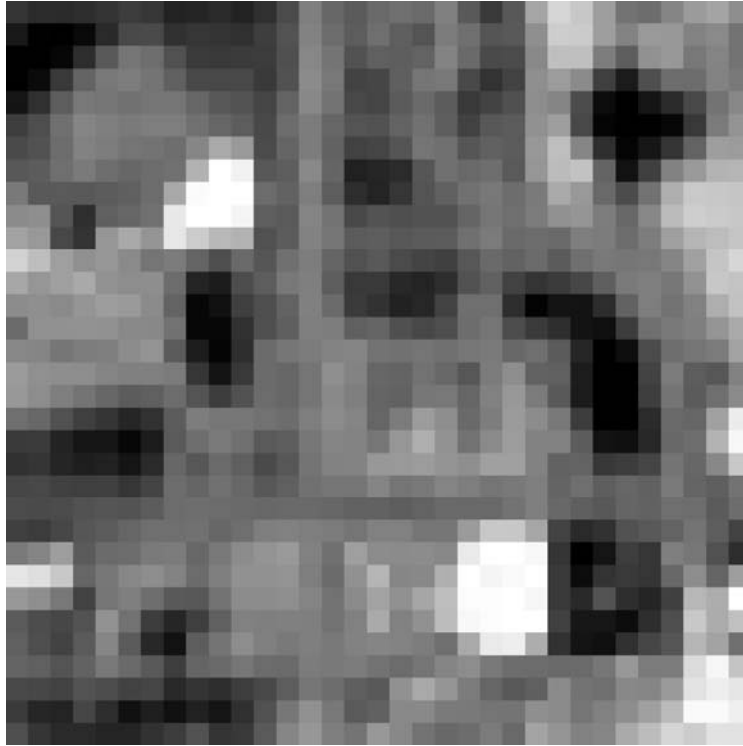
**Figure 1-5.** a) PAN IKONOS image of Columbia (1 m resolution), b) false color infrared IKONOS MS image of Columbia (4 m resolution), and c) false color infrared PS-MS IKONOS image of Columbia (1 m resolution).

of its simplicity and the high quality results of the sharpening, the color normalization method is used to produce PS-MS images from the IKONOS datasets analyzed in this

research. An example of the construction of a PS-MS image from IKONOS data of Columbia, MO is shown in Fig. 1-5.

#### *E. Urban Area Imagery*

Due to varying spatial and spectral characteristics, the appearance of urban areas in images from airborne and satellite-based sources varies significantly. The spatial resolution, spectral content, and noise levels of the different image sources impacts their utility for urban area analyses. Figures 1-6 through 1-9 show images of a commercial area from Landsat ETM+ PAN, DOQQ, NIMA/NGA, and IKONOS data sets, respectively. It is clear from Fig. 1-6 that the 15-m Landsat image is of little use for detailed urban analysis. While the DOQQ (Fig. 1-7) and IKONOS PAN (Fig. 1-9a) images both have a 1-m spatial resolution it is clear that the increased dynamic range and decreased noise levels in the IKONOS imagery make it much more suitable for image processing and analysis. The fine spatial resolution of the NIMA/NGA imagery (Fig. 1-8) makes it possible to clearly identify very small features, such as automobiles, in the imagery. In addition, the presence of the RGB spectral bands allow for further analysis such as classification to be performed. The NIR band in the IKONOS MS data (Fig. 1-9b) greatly facilitates the automatic identification of vegetation in the image (see Section 4.4-A). Finally, as was discussed in sub-section 1.2-D above the MS and PAN bands can be fused to produce a 1-m resolution PS-MS image. The IKONOS PS-MS imagery (Fig. 1-9c) have distinct advantages over all other image types discussed thus far because they have the combined benefits of high-spatial resolution, 11-bit per pixel dynamic range, and 4-band multi-spectral content.



**Figure 1-6.** Landsat ETM+ PAN image of commercial land use area (15 m resolution).



**Figure 1-7.** DOQQ orthophoto of commercial land use area (1 m resolution).



**Figure 1-8.** NIMA orthophoto of commercial land use area (0.3 m resolution).



**Figure 1-9a.** IKONOS PAN image of commercial land use area. Note the difference in image quality between the IKONOS data and the DOQQ image (both 1 m resolution).





**Figure 1-9b.** False color IKONOS MS image of commercial land use area (4 m resolution).



**Figure 1-9c.** False color IKONOS PS-MS image of commercial land use area (1 m resolution).

### **1.3. Satellite Image Analysis and Classification**

A large number of image processing and classification techniques have been developed and successfully applied to remotely sensed imagery for a wide variety of applications. Image processing techniques generally seek to either suppress or remove noise from the imagery, or to accentuate certain characteristics in the imagery to ease interpretation or further analysis of the imagery. The description of techniques to accentuate specific image characteristics will make up the bulk of the image processing techniques reviewed here. Classification techniques seek to classify objects, in this case image pixels or features extracted from them, into a number of distinct categories or classes. A variety of classification techniques commonly used with remote sensing data are reviewed. A common problem encountered in remote sensing image analysis is that of mixed pixels, i.e. the presence of more than one land cover type within a single image pixel. One approach for dealing with this problem is to use fuzzy classification techniques, where each pixel can be classified as more than one land cover type. A brief review of fuzzy set theory and fuzzy classification is also presented.

#### *A. Texture Analysis and Smoothing Filters*

An important and easily visible attribute of remotely sensed images is that they are highly textured. The amount and properties of the texture varies significantly throughout the imagery, depending on land cover type. Therefore, texture features extracted from the imagery can be used to help discriminate between various land cover types. This is especially useful when analyzing images from urban areas where many of the land cover types exhibit similar spectral signatures. Texture can be defined as the

spatial distribution of tonal variations within a band of data [Haralick, 1973]. Texture features measure properties such as smoothness, coarseness, regularity, and randomness in the image. While there are a variety of techniques for the extraction of texture features, including first and second order statistical approaches, Fourier spectrum techniques, and wavelet-based approaches, statistical texture features are the simplest and most common. Statistical texture features are calculated from the pixel values within a window surrounding the pixel of interest. When statistical texture features are calculated, the output is a new image with the same dimensions as the input image where the digital pixel values represent the strength of the texture feature being measured. These transformed texture images can then be used for further analysis or as input into a classification algorithm.

First order statistical texture measures are calculated from the normalized gray-level histogram,  $p(z_i)$ , of the pixel window  $w(z_i)$ , where  $0 \leq i \leq L-1$ ,  $L$  is the number of gray levels in the image, and  $z$  is a random variable denoting image gray levels [Gonzalez and Woods, 2002]. These first-order texture measures are calculated as follows:

$$entropy = -\sum_{i=0}^{L-1} p(z_i) \log_2 [p(z_i)] \quad (1-2)$$

$$data\ range = \max[w(z_i)] - \min[w(z_i)] \quad (1-3)$$

$$variance = \sigma^2 = \sum_{i=0}^{L-1} (z_i - \mu)^2 p(z_i) \quad (1-4)$$

$$skewness = \frac{1}{\sigma^3} \sum_{i=0}^{L-1} (z_i - \mu)^3 p(z_i) \quad (1-5)$$

where  $\mu$  is the mean value of the gray levels in the window, i.e.



$$\mu = \sum_{i=0}^{L-1} z_i p(z_i) \quad (1-6)$$

The size of the pixel window,  $w(z_i)$ , is variable and has a significant impact on the calculation of the texture features. A larger pixel window allows for a better quantification of the texture present in a region because there are more samples for the calculation. However, larger windows lead to an increase in the “edge effect” where erroneous texture values are calculated along the edges between regions with differing textures. This problem is considered in depth in the work of Ferro and Warner [2002].

Noise suppression or image smoothing is an important image-processing task in many applications. As was the case with the statistical texture measures discussed above, smoothing filters are implemented by passing a window, or kernel, over the image. The value of each pixel in the output image is a function of the pixels in the kernel. The amount of smoothing is determined both by the type of filter employed and by the size of the kernel. Larger kernels remove more noise, however they may also result in significant blurring of the image. The simplest smoothing filter is to simply take the average of the pixels in the kernel. More advanced smoothing filters such as Butterworth and Gaussian lowpass filters can also be used.

A smoothing filter that is particularly useful in many situations is the median filter. The median filter is a type of order-statistics filter. Order-statistics filters are nonlinear filters where the response is based on the sorted order of the pixels contained in the kernel. As the name suggests, the output of the median filter is the median value of the pixels in the kernel. The unique property of the median filter is that it preserves edges in the image, resulting in significantly less blurring. The median filter effectively removes objects from the image that are smaller than the kernel. This is particularly

useful when processing high-resolution images of urban areas because it can be used to remove small objects such as automobiles from the image. For a more detailed description of the smoothing filters reviewed here see Gonzalez and Woods [2002].

### *B. Classification Techniques*

Classification techniques can be divided into supervised and unsupervised approaches. Supervised classification approaches seek to exploit *a priori* knowledge about the data to identify which of the  $M$  predefined classes,  $\omega_1, \dots, \omega_M$ , each data point belongs. This *a priori* knowledge typically takes the form of training data, a collection of data points for which the class labels are known. The training data are then used by the classifier to identify the class labels of the unknown data points. However, there are many situations in which there are no training data available, or even the classes present in the data are not known, in which case unsupervised approaches must be used. Unsupervised classification techniques seek to identify the underlying divisions in the data and then group similar data points together into clusters. For a more detailed description of a variety of classification techniques see Jenson [1996] and Theodoridis and Koutroumbas [1999].

Three of the most commonly used supervised classification algorithms for analysis of MS remotely sensed imagery are the parallelepiped, minimum distance to means, and maximum likelihood classifiers. Both the parallelepiped and minimum distance to means algorithms are non-parametric (i.e. no statistical distribution of the data points is assumed), whereas the maximum likelihood classification algorithm is parametric. Before the classification algorithms are described, it is necessary to define

the notation that will be used. Here we define the vector  $\mathbf{x}$  as an MS data point, and  $\mathbf{x}^c$  is a training data point from class  $c$ , where  $c = 1, \dots, M$ , and  $M$  is the number of classes.  $x_i$  is the scalar value of band  $i$  in sample  $\mathbf{x}$ ,  $i = 1, \dots, b$ , where  $b$  is the number of bands in the MS data.  $\boldsymbol{\mu}^c = [\mu_1^c, \dots, \mu_b^c]^T$  is the mean vector of class  $c$  and  $\mu_i^c$  is the mean of band  $i$  of the training data from class  $c$ , and  $N_c$  is the number of training samples from class  $c$ .

The parallelepiped classification algorithm is based on a simple Boolean logic decision rule. First, the mean vector,  $\boldsymbol{\mu}^c$ , and standard deviations,  $\sigma_i^c$ , are calculated for each class from the training data.

$$\mu_i^c = \frac{1}{N_c} \sum_{j=1}^{N_c} x_{i,j}^c \quad (1-7)$$

$$\sigma_i^c = \left( \frac{1}{N_c} \sum_{j=1}^{N_c} (x_{i,j}^c - \mu_i^c)^2 \right)^{\frac{1}{2}} \quad (1-8)$$

where  $x_{i,j}^c$  is the value of band  $i$  of the  $j^{\text{th}}$  sample of the training data from class  $c$ . A standard deviation threshold,  $s_\sigma$ , is set by the user. Sample,  $\mathbf{x}$ , is classified as  $\omega_c$  if and only if

$$\mu_i^c - s_\sigma \cdot \sigma_i^c \leq x_i \leq \mu_i^c + s_\sigma \cdot \sigma_i^c \quad \forall i = 1, \dots, b. \quad (1-9)$$

In effect, the classifier defines  $M$   $b$ -dimensional parallelepipeds around each of the mean vectors, the size of which is defined by  $s_\sigma$ . Each sample is classified as the class of the parallelepiped into which it falls. If the sample does not fall into any of the parallelepipeds, it is labeled as unclassified.

The minimum distance from means classification algorithm classifies samples by determining the closest class mean to the sample. There are a variety of distance

measures that can be used, but the most common is the Euclidean distance,  $d_e(\mathbf{x}, \mathbf{y})$ , calculated as

$$d_e(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \sqrt{\sum_{i=1}^b (x_i - y_i)^2} \quad (1-10)$$

Or, in vector notation as

$$d_e(\mathbf{x}, \mathbf{y}) = ((\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}))^{1/2} \quad (1-11)$$

For each sample,  $\mathbf{x}$ , the distance to each of the class means,  $\boldsymbol{\mu}^c$ , is calculated and the sample point is classified as belonging to the class to which it is the closest. The sample  $\mathbf{x}$  is classified as class  $\omega_c$  if

$$d_e(\mathbf{x}, \boldsymbol{\mu}^c) < d_e(\mathbf{x}, \boldsymbol{\mu}^k) \quad \forall k \neq c \quad (1-12)$$

Unlike the parallelepiped classifier, all data samples are assigned a class, and the minimum distance to means classification algorithm does not utilize the variance of the training data during the classification process.

The most commonly used classification algorithm for remotely sensed imagery is the maximum likelihood classifier. Maximum likelihood classification assumes that the training data statistics are Gaussian distributed, which for many data sources is a valid assumption. This classification scheme is based on Bayes decision theory, and makes full use of both the mean and covariance statistics from the training data. The Bayes classification rule is to classify sample  $\mathbf{x}$  as the class that has the maximum *a posteriori* probability given the sample,  $P(\omega_k | \mathbf{x})$ . So,  $\mathbf{x}$  is classified as  $\omega_c$  if and only if

$$P(\omega_c | \mathbf{x}) > P(\omega_k | \mathbf{x}) \quad \forall k \neq c \quad (1-13)$$

The *a posteriori* probability,  $P(\omega_c | \mathbf{x})$ , is calculated using Bayes rule

$$P(\omega_c | \bar{\mathbf{x}}) = \frac{p(\bar{\mathbf{x}} | \omega_c)P(\omega_c)}{p(\bar{\mathbf{x}})} \quad (1-14)$$

where  $p(\mathbf{x}|\omega_c)$  is the class-conditional probability density function (pdf),  $P(\omega_c)$  is the *a priori* probability of class  $\omega_c$ , and  $p(\mathbf{x})$  is the pdf of  $\mathbf{x}$ , calculated using

$$p(\bar{\mathbf{x}}) = \sum_{k=1}^M p(\bar{\mathbf{x}} | \omega_k)P(\omega_k) \quad (1-15)$$

Assuming the data follows the multivariate normal density distribution

$$p(\bar{\mathbf{x}} | \omega_c) = \frac{1}{(2\pi)^{b/2} |\Sigma_c|^{1/2}} \exp\left(-\frac{1}{2}(\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}^c) \Sigma_c^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}^c)\right) \quad (1-16)$$

where  $\bar{\boldsymbol{\mu}}^c$  is calculated using Eq. (1-7) and  $\Sigma_c$  is the covariance matrix of the training data from class  $c$ , calculated as

$$\Sigma_c = \frac{1}{N_c} \sum_{j=1}^{N_c} (\bar{\mathbf{x}}_j^c - \bar{\boldsymbol{\mu}}^c)(\bar{\mathbf{x}}_j^c - \bar{\boldsymbol{\mu}}^c)^T \quad (1-17)$$

Eq. (1-16) can be simplified to ease the computational burden, deriving the likelihood function for class  $\omega_c$ ,  $g_c(\mathbf{x})$

$$g_c(\bar{\mathbf{x}}) = \ln(P(\omega_c)) - \frac{1}{2} \ln(|\Sigma_c|) - \frac{1}{2} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}^c)^T \Sigma_c^{-1} (\bar{\mathbf{x}} - \bar{\boldsymbol{\mu}}^c) \quad (1-18)$$

The classification rule is the same as Eq. (1-13), except the likelihood functions for each class are compared instead of the *a posteriori* probabilities. Often times, the *a priori* class probabilities are not available during classification, in which case the  $\ln(P(\omega_c))$  term would be left out of Eq. (1-18). Because maximum likelihood classification is such a commonly used technique for classification of remotely sensed imagery, it is used as a benchmark with which to compare other classification techniques developed in this research. Results of maximum likelihood classification of IKONOS data for urban land cover mapping are presented in Chapter 2.

Two commonly used techniques for unsupervised classification of remotely sensed imagery are clustering and segmentation. Clustering techniques seek to identify natural divisions in the data and group the samples into clusters based on a similar spectral response or similar values of features extracted from the data. Segmentation is very similar to clustering, with the added constraint of connectivity in the image. Segmentation techniques attempt to subdivide an image into regions or objects consisting of similar pixels. Both clustering and segmentation are iterative techniques where the clusters/segments are formed through a series of repeated steps until convergence is reached.

There are many different types of clustering algorithms and we refer the reader to Theodoridis and Koutroumbas [1999] for a more detailed discussion. Here, we present a review of the  $k$ -means clustering technique because of its widespread use and simplicity. Before the algorithm begins, the number of clusters to be identified in the dataset,  $k$ , is selected by the user. Initially, the mean vectors,  $\mu^1, \dots, \mu^k$  are randomly distributed throughout the dataset. For each sample  $x$ , the distance to each of the cluster mean vectors is calculated, and  $x$  is classified as belonging to the cluster to whose mean vector it is closest. The mean vectors are then recalculated based on the data points now assigned to each cluster. The classification of the data samples and recalculation of the cluster mean vectors is repeated until there is no change in the value of the cluster mean vectors in two successive iterations. While Euclidean distance is the most common distance measure used, other distance measures such as the Mahalanobis distance, which takes into account the covariance matrix can be used. The Mahalanobis distance,  $d_m$ , is defined as

$$d_m(\mathbf{x}, \boldsymbol{\mu}_c) = ((\mathbf{x} - \boldsymbol{\mu}^c)^T \boldsymbol{\Sigma}_c^{-1} (\mathbf{x} - \boldsymbol{\mu}^c))^{1/2} \quad (1-19)$$

If  $d_m$  is used for the classification, the covariance matrices for each cluster must also be recalculated after each iteration.

As mentioned above, segmentation is an unsupervised classification technique that seeks to subdivide an image into constituent regions or objects. There are many approaches to segmentation such as thresholding, edge detection, region growing/merging, and morphological watersheds. For an extended discussion of segmentation techniques and their application to different types of images we refer the reader to Gonzalez and Woods [2002]. A variety of segmentation techniques have been applied to remote sensing imagery with varying degrees of success. Segmentation of remotely sensed images is a difficult problem due to mixed pixels, spectral similarity, and the textured appearance of many land-cover types. Many segmentation algorithms are based on a region growing/merging approach where pixels are iteratively grouped into regions based on predefined similarity criteria. Region growing techniques begin with ‘seed’ pixels that meet some criteria and then from these points allow regions to grow by adding pixels that meet a similarity criterion. Region merging techniques begin with every pixel in the image defined as a segment and segments are merged together which meet a similarity criterion. Both of these techniques require the definition of the similarity rules that the constituents of a segment must meet and the definition of a stopping criterion to determine when the algorithm should stop the growing/merging process. A detailed description of the region merging technique used in this research is presented in Chapter 3.

### *C. Fuzzy Set Theory and Classification*

The classification techniques reviewed above all produce crisp classifications, i.e. each pixel is assigned to only one class. However, remote sensing images are imprecise in nature since they contain mixed pixels and many land cover types with similar/overlapping spectral signatures. These problems are particularly severe in urban environments where much of the land cover is composed of roads, buildings, and other types of impervious surface that often have similar spectral signatures. The urban environment also contains many small sub-pixel features leading to large numbers of mixed pixels. Although high-resolution sensors help to alleviate the problem of mixed pixels, they are still present to some extent. Fuzzy classification techniques allow pixels to have membership in more than one class and therefore better represent the imprecise nature of the data Wang [1990].

Fuzzy classification approaches are based on fuzzy set theory. Here we present a very brief overview of fuzzy set theory and fuzzy systems; for a detailed description see Klir and Yuan [1995]. Let  $X$  represent the universal set with elements  $x$ , i.e.  $X = \{x\}$ . In classical set theory elements are viewed as either belonging entirely to a set or not at all; sets of this type are referred to as crisp sets. Membership of an element  $x$  in crisp set  $A$  can be represented by a characteristic function,  $\chi_A$ , such that  $\chi_A(x) = 1$  if  $x \in A$  and  $\chi_A(x) = 0$  if  $x \notin A$ . No other values of set membership are possible in classical set theory. However, elements in fuzzy set theory can take on partial membership in a set, allowing for the representation of imprecise or vague data. The characteristic function  $\chi_A$  of classical set theory is generalized to the membership function  $f_A$  of fuzzy set theory, such that  $f_A(x)$  is equal to a real number in  $[0,1]$  indicating the degree of membership of



element  $x$  in fuzzy set  $A$ . Values of  $f_A(x)$  close to 1 indicate a high degree of membership of  $x$  in  $A$ .

The three basic operations used for the manipulation and combination of crisp sets are union, intersection, and complement. The union of two sets,  $A \cup B$ , consists of all elements  $x$  that belong to either  $A$  or  $B$ . The intersection of sets  $A$  and  $B$ ,  $A \cap B$ , consists of the elements  $x$  that belong to both  $A$  and  $B$ . The complement of set  $A$ ,  $\bar{A}$ , consists of all elements  $x$  that are not in  $A$ . These operations all have extensions to fuzzy set theory. While there are many ways to generalize these set operations to fuzzy sets, one group is used most often and referred to as standard fuzzy set operations. They are the standard fuzzy union, standard fuzzy intersection, and standard fuzzy complement:

$$f_{A \cup B}(x) = \max[f_A(x), f_B(x)] \quad (1-20)$$

$$f_{A \cap B}(x) = \min[f_A(x), f_B(x)] \quad (1-21)$$

$$f_{\bar{A}}(x) = 1 - f_A(x) \quad (1-22)$$

These operators can be combined into fuzzy logic rules to perform reasoning operations on imprecise data in a fuzzy system. When utilized in this fashion, the fuzzy union, intersection, and complement operators are referred to as OR, AND and NOT, respectively. Fuzzy systems have been applied to a wide variety of problems including both supervised and unsupervised classification. The outputs of a fuzzy logic rule base are fuzzy membership values, however for most classification applications a crisp value is desired. For example, the desired output of a classification of a remote sensing image is a land cover map, which requires a crisp value for each pixel. The process of converting fuzzy membership values to crisp values is called defuzzification. Even though the final output of many fuzzy classifiers is a crisp output, there is still value in

using a fuzzy system for classification. This approach allows the imprecision inherent in remote sensing imagery to be accounted for during the classification process. The final classification of a pixel is not made until the very last step in the system, allowing different pieces of evidence from different sources to be evaluated to determine the class label. This is a distinct advantage in a classification framework that seeks to utilize both spectral and spatial information (e.g. texture) to discriminate between various land cover classes.

#### **1.4 Summary of Related Work**

In this section, we review previous work done in the areas of analysis and classification of remotely sensed imagery. The works reviewed here are subdivided into four categories. First, work pertaining to general techniques of classification and segmentation of a variety of remote sensing data sources and are presented. Second, work that has been done specifically on analysis of high-resolution imagery from urban areas is reviewed. Next, a review of research on the extraction of roads from remotely sensed imagery is presented. Finally, a review of research on the extraction of buildings from remotely sensed imagery is presented.

##### *A. Fuzzy Classification and Segmentation Techniques*

A supervised fuzzy classifier was introduced by Wang [1990] and applied to Landsat MSS and TM data. Wang's classifier used the maximum likelihood algorithm in conjunction with a fuzzy mean vector and fuzzy covariance matrix. This classification technique was able to identify and determine the content of mixed pixels, and also

resulted in an overall accuracy increase when compared to traditional maximum likelihood classification. Another supervised fuzzy classification approach was proposed by Melangi *et al.* [2000] and applied to Landsat TM data. Membership functions are represented by normalized Gaussians, and classification is performed using a MIN fuzzy reasoning rule. While this technique was able to identify and determine the content of mixed pixels, the overall classification accuracy was less than the traditional maximum likelihood classification. Bardossy and Smaniego [2002] developed a fuzzy logic rule-based classifier and applied it to Landsat TM data. Triangular fuzzy sets were used, and the rule base was generated via the simulated annealing optimization technique. This technique produced a slightly more accurate classification than the maximum likelihood algorithm, and like the other fuzzy classifications techniques reviewed, mixed pixels were successfully identified. In addition, a fuzzy rule base was generated that could be analyzed to better understand the unique properties of each land cover class.

A number of segmentation techniques have been developed and applied to remote sensing imagery with varying degrees of success. A segmentation technique based on a combination of region growing and spectral clustering was presented by Tilton [1998] and applied to Landsat MSS data. This approach alternates between region growing and spectral clustering in each iteration and produces a pre-specified number of hierarchical segmentation results. The hierarchy of segmentation results ranges from very coarse, with only several large segments, to fine where small details in the image are distinguished. Baatz and Schape [2000] developed a region merging approach for multi-scale hierarchical segmentation based on local homogeneity criteria that take into account both spectral and shape information. This technique was able to successfully segment

textured regions and each individual segmentation in the hierarchy consists of image objects of similar size. The algorithm was tested on several different data sources, and positive results were reported for each. This approach is described in greater detail in Chapter 3, and results of its application to IKONOS imagery are presented.

A segmentation approach for Landsat TM data was presented in the work by Evans *et al.* [2002]. This technique is a hybrid boundary/region-based approach where seeds for region growing are extracted from the MS imagery with a local canonical transformation that identifies region boundaries. This technique was tested for imagery from agricultural regions and successfully discriminated fields with little over or under-segmentation. A combined segmentation and classification approach for MS imagery based on a Markov random field model was presented by Sarkar *et al.* [2002]. The image is initially over-segmented with a standard segmentation technique, and these segments then are merged using a Markov random field optimization approach. Training data is then used to identify the class labels for the image objects identified during the segmentation step. The proposed combined segmentation and classification approach significantly outperformed maximum likelihood classification when tested on data from the IRS-1B and IRS-1C Indian satellites. A similar combined approach is described in this research in Chapter 3.

### *B. Analysis of High-Resolution Imagery from Urban Areas*

Pesaresi [1999] conducted a study on the use of texture features for urban-area classification. 5-m spatial resolution PAN imagery from the IRS-1C Indian satellite was analyzed to determine the relationship between classification accuracy and the window

size used for the calculation of texture features. The land cover classes were general urban classes such as 'built-up', and were classified using maximum likelihood classification. Large window sizes of 91 and 181 pixel squares were found to give the highest classification accuracy. In the work by Tatem *et al.* [2001] a Hopfield neural network was used to perform sub-pixel classification of MS IKONOS imagery. In this approach, a fuzzy classification was first performed to identify mixed pixels. The Hopfield neural network then used the class proportions identified in the fuzzy classification to produce a classification at increased resolution by dividing the mixed pixels into multiple classes. The results of applying a contextual classification technique to MS IKONOS imagery were presented by Van Teeffelen *et al.* [2001]. The imagery is initially classified into detailed urban land cover classes before being analyzed by the contextual filtering algorithm, which reclassified the image into land use categories based on spatial context of the classified pixels. This technique was able to successfully identify different land use types such as industrial sites, commercial areas, and new settlements in the imagery.

An object-oriented classification approach was applied to MS and PAN IKONOS data by Kressler *et al.* [2001]. Initially, the image is segmented with a multi-resolution segmentation technique. The image objects are subsequently classified with a fuzzy rule-base classifier. It was found that this technique was able to more accurately classify large objects than small objects in the imagery. A technique for building extraction from IKONOS imagery was introduced by Lee *et al.* [2003]. Supervised classification was performed to find initial building candidates, followed by segmentation and squaring based on the Hough transform for the building extraction. Using this technique,

approximately 64% of the buildings present in the test imagery were extracted. Pesaresi and Benediktsson [2001] developed a segmentation approach for high-resolution imagery based on geodesic morphological operators. A differential morphological profile is constructed from geodesic opening and closing images calculated with an increasing structuring element size. The differential morphological profile is used to characterize the size and contrast of homogenous regions in the imagery. This technique was tested on several high-resolution datasets of urban areas and found to work well in complex scenes where other segmentation techniques have difficulty. A neural network classification approach based on the differential morphological profile was presented by Benediktsson *et al.* [2003]. Here, urban areas are characterized by a differential morphological profile. A decision boundary feature extraction algorithm was used to reduce the dimensionality of the profile before classification was performed by a conjugate gradient neural network. This classification technique was able to classify detailed urban land cover classes with a high degree of accuracy.

### *C. Road Extraction*

A technique for the detection of curvilinear structures was presented by Steger [1998]. Detection of line points in the imagery was accomplished by convolving the image with discrete two-dimensional partial derivative kernels. A linking algorithm then groups these line points together. After the initial line detection has been completed, the width of the lines is determined. This technique was applied to aerial photography and obtained good results for roads in rural areas. Couloigner and Ranchin [2000] introduce a semi-automated approach for road extraction that makes use of multi-resolution

analysis and a wavelet transform. Here, the user provides start and end points for each road, and the algorithm then extracts road edges and topology. This method successfully extracts urban streets laid out in a rectangular grid. A technique to extract road networks from a classification of high-resolution IKONOS urban imagery was presented by Shi and Zhu [2002]. A 'line segment match method' was used to detect long linear groups of pixels classified in the road class. These road pixels are then simplified into the road centerlines through the use of morphological operators. Accuracies of approximately 90% are reported for a small test site. In their review of mathematical morphology applications for remote sensing imagery, Soille and Pesaresi [2002] present an example of extracting roads with directional morphological profiles. The directional profiles are found by computing morphological openings and closings with linear structuring elements of varying orientations. These measurements are used to determine the orientation of the roads.

A road extraction technique based on a rotating texture filter was described by Haverkamp [2002]. Potential road points were selected by identifying the directions of the minimum response obtained from the texture filter. These points are then joined together to construct the road network through the use of a network model. Road extraction accuracies of 49% to 66% were reported. Hinz and Baumgartner [2003] presented a complex road network extraction approach that attempts to accurately map both the road network and the road edges through the use of snakes. This approach utilized multiple very high-resolution aerial images, detailed scene models, and error self-diagnosis to perform the road extraction. This technique results in a road extraction accuracy of 95%. Doucette *et al.* [2001] introduced a self-organizing road map algorithm

to extract roads from high-resolution MS imagery. The self-organizing road map, a specialized version of the self-organizing map neural network, performs spatial clustering to identify and group together elongated regions. The algorithm operates on a supervised classification image. In addition to optical imagery, road network extraction can also be performed on synthetic aperture radar (SAR) data. An example of this was presented by Dell'Acqua and Gamba [2001] where fuzzy clustering algorithms were used for road extraction. After the data were classified with an unsupervised fuzzy clustering algorithm, roads were extracted via fuzzy shortest path extraction.

#### *D. Building Extraction*

A technique for building extraction from supervised classification of IKONOS imagery is presented by Lee *et al.* [2003]. First, a supervised classification of the MS data is performed, focusing on identification of different types of building roof materials. The classification is then used to define a window within the PAN imagery within which to search for buildings. Buildings are then extracted using unsupervised clustering followed by a polygon squaring technique based on the Hough transform. Building detection rates of 64%-72% and quality values of 30%-51% are reported. Croitoru and Doytsher [2003] presented a model-based building extraction technique that relies on the detection of right-angle building corners. Buildings are detected using pose clustering, a voting technique, where right-angle corners are used as voting elements. The voting process is constrained by detection of shadowed regions. This approach is limited by the necessity of a model for every building shape that is to be detected. Building detection rates of 79%-81% and quality values of 32%-34% are reported. A technique for



detecting small objects, such as house plots and vegetation patterns, in 4.5-m PAN satellite imagery based on combining supervised shape classification with unsupervised image segmentation was presented by Segal and Kaufmann [2001]. This approach utilizes a thresholding segmentation technique that has been modified to identify specific shapes on which it has been trained. Extraction completeness values of 61%-97% and correctness values of 58%-94% are reported. Jaynes *et al.* [1994] present a task driven perceptual organization approach for extraction of building polygons. This approach has three steps: low-level feature extraction of line segments and orthogonal corners, perceptual grouping of those features into chains, and polygon hypothesis generation from the chains. The building extraction algorithm was tested on high-resolution aerial photography and resulted in the detection of 87% of the buildings, however the accuracy of the extracted buildings was not reported. A technique based on extracting buildings from high-resolution PAN aerial photography by first identifying their cast shadows was developed by Irvin and McKeown [1989]. Building hypotheses are initially generated by grouping detected edges and corners into polygons. Shaded regions are then identified using a variety of image processing techniques, followed by identification and simplification of the shadow edge. The shadow edges are then combined with the building hypotheses to identify buildings. Building extraction accuracy measures for this approach were not reported.

## **1.5 Summary and Research Objectives**

In this chapter a variety of background material was presented on geospatial information products, high-resolution satellite imagery, and image processing techniques

and methods. The utility of these for urban area applications was discussed. Geospatial information products developed from commercial high-resolution satellite imagery have numerous government applications at both the local and national level. For example, the USGS's National Map integrates a variety of geospatial information products (topography, roads, land cover, etc.). In addition to civilian government applications, there are many defense and intelligence applications for urban area geospatial information products. NIMA/NGA has integrated the previously disparate areas of mapping and imagery to produce a new intelligence tradecraft called geospatial intelligence or GeoINT.

Several earth image collection methods were discussed, including both aerial photography and satellite based digital sensors. In particular, high-resolution imagery from commercial satellites such as IKONOS and QuickBird are well suited for use in many urban area applications. Following the discussion of the image sensor systems, a number of relevant image analysis techniques were reviewed. A brief review of texture analysis, focusing on first-order texture measures, and image smoothing techniques was presented. Both supervised and unsupervised classification techniques were reviewed, and several techniques, such as maximum likelihood, were discussed in detail. This was followed by an introduction to fuzzy set theory, fuzzy logic, and their application to remote sensing image analysis. Finally, a summary of related work in the areas of classification techniques, segmentation techniques, analysis of urban areas, automated road extraction, and automated building extraction was presented.

Section 1.1 discussed in detail the numerous applications for urban area geospatial information products. Furthermore, commercially available high-resolution satellite

imagery is an ideal data source from which these information products can be derived. Due to the fact that commercial high-resolution image data have only recently become available, relatively little work has been done on their exploitation for urban applications. The goal of this doctoral research is to develop advanced automated and semi-automated image analysis and classification techniques for the extraction of urban area geospatial information products from commercial high-resolution satellite imagery. Specifically, the problems of both semi and fully automated urban land cover classification and automated extraction of urban infrastructure features, such as road network and 2D building vector features, will be addressed. The development of automated and semi-automated techniques for generation of these urban geospatial information products is of high importance not only because of the many applications for which they can be used but also because the large volume of data collected by these sensors exceeds the human capacity of trained image specialists to analyze. Many applications, especially military and intelligence related applications, require near real time exploitation of image data. Both the quantity of data and the speed with which that data must be analyzed illustrate the need for automated geospatial information extraction. While some of the techniques developed in this research are only semi-automated, requiring the user to input training data, they still speed up the process of extracting information from satellite imagery.

The practicality and usefulness of geospatial information products generated via automated techniques is dependent on both the quality of the extracted information and the application in which they are being utilized. Geospatial information products are currently extracted manually by trained image analysts, and are therefore highly accurate. There is no information reported in the literature on how accurate automatically extracted

geospatial information products must be to be usable. The point at which the tradeoff between information accuracy and number of man-hours required for data production becomes unprofitable is unknown. Although the required accuracy values cannot be quantified, it is clear that the level of accuracy required for automatically extracted geospatial information products is application dependent. For example, emergency response and management applications would require highly accurate information as emergency responders on the ground might be using the information for navigational purposes, or to determine the areas in most need of assistance. Other applications, such as urban growth modeling, can utilize extracted features with lower levels of accuracy and still be successful. One of the most promising applications of automated techniques is in assisting image analysts in the feature extraction process. Here, the goal is to produce a highly accurate, but not necessarily complete extraction of the feature of interest. For example, if an automated building extraction technique is able to accurately identify 75% of the buildings present in the image being analyzed, then the image analyst only has to manually digitize 25% of the buildings, greatly reducing the time required for data production. Another application where automated extraction techniques are useful even if they are not completely accurate is information mining of remote sensing image database archives. Here, the goal is to enable a user to query the database for images containing certain land cover types or features. Image archives containing feature or content information are needed to catalog the vast quantities of data being generated by satellite sensors. Automated extraction techniques can be used to characterize the image content as it is acquired, increasing the accuracy of user queries.

The rest of this document is organized as follows. In Chapter 2, a hierarchical fuzzy pixel-based classification scheme is introduced which examines both spectral and spatial information during the classification process and results in a significant increase in classification accuracy when compared to maximum likelihood. A combined pixel-based and object-based approach for urban land cover classification is presented in Chapter 3. The results of the fuzzy pixel-based classification are combined with segmentation and shape analysis to produce a more detailed urban land cover map than was possible with the pixel-based approach. In Chapter 4, the problem of automated road network extraction is addressed. Two techniques for extracting roads from a land cover classification image are presented, followed by a fully automated technique for extracting the road network directly from the imagery. In Chapter 5, a fully automated multi-detector building footprint extraction technique is presented. Next, a fully automated urban land cover classification approach is presented in Chapter 6. This technique utilizes fully automated feature extraction techniques for the generation of training data, which are then fed into the classification approaches presented in Chapters 2 and 3. Finally, concluding remarks as well as future directions for this research are discussed in Chapter 7.

## Chapter 2

### Fuzzy Pixel-Based Urban Land Cover Classification

#### 2.1 Introduction

As discussed in Chapter 1, there are a wide variety of applications for geospatial information products derived from high-resolution multispectral satellite imagery over urban areas. One very important geospatial information product that can be extracted from this type of imagery is the land cover map, which can be generated using supervised classification of the multispectral image data. Because of the complex nature and diverse composition of land cover types found within the urban environment, classification of high-resolution multispectral satellite imagery is a difficult task. Materials typically found in the urban environment include concrete, asphalt, metal, plastic, glass, shingles, water, grass, trees, shrubs, and soil, to list just a few. Moreover, many of these materials are spectrally similar, and this leads to problems in automated or semi-automated image classification of these areas when using multispectral data comprised of only four spectral bands. In addition, these materials form very complex arrangements in the imagery such as housing developments, transportation networks, industrial facilities, and commercial/recreational areas.

Conventional methods for classification of multispectral (MS) remote sensing imagery such as parallelepiped, minimum distance from means, and maximum likelihood (see Chapter 1, Section 1.3-A) only utilize spectral information and consequently have

limited success in classifying high-resolution urban MS images. As many classes of interest in the urban environment have similar spectral signatures, spatial information such as texture and context must be exploited to produce accurate urban land cover maps from supervised or unsupervised classification techniques. The exploitation and use of spatial information is only possible because the high spatial resolution (e.g. 1 m) commercial satellite imagery has only recently become available (Chapter 1). As a result, very little research has been done that incorporates both spectral and spatial information in land cover classification schemes.

In this chapter, we first present the results of a study done to analyze the suitability of the traditional maximum likelihood classification approach for production of detailed high-resolution urban land cover maps. Then, the use of several spatial features is investigated to determine their effect on the discrimination at the pixel level between spectrally similar urban land cover classes. A hierarchical fuzzy classification technique is then presented that alleviates some of the shortcomings of conventional classification techniques by incorporating the spatial features along with the spectral information. Finally, the results of applying the hierarchical fuzzy classifier to images from several urban test sites are presented.

## **2.2 Maximum Likelihood Classification Study**

The effectiveness of traditional maximum likelihood classifiers applied to high-resolution satellite imagery of urban and suburban scenes was assessed to provide a baseline for comparison with classification techniques developed in this research. The imagery used for this study was acquired by the IKONOS commercial remote sensing

satellite. Two IKONOS image datasets are used in this study: an image of Columbia, Missouri acquired on April 30, 2000, and an image of Springfield, Missouri acquired on September 17, 2000. Both image datasets include a variety of urban and suburban land cover types making them ideal for this study. Two separate datasets were used to provide multiple evaluations of the classification algorithms developed in our research and to ensure that the algorithms were not so highly specialized as to be applicable to only a single dataset.

The panchromatic (PAN) band of the Columbia IKONOS image is shown in Fig. 2-1. The IKONOS images went through several preprocessing steps before classification. First, the images were orthorectified to increase the planar accuracy from 25 m RMS to approximately 3 m RMS [Davis and Wang, 2003]. Map-quality positional accuracy is needed so that the image data and derivative products (e.g. land cover map) can be effectively incorporated into GIS databases. After orthorectification, the color normalization method (Eq. 1-1) was used to fuse the PAN data with the four MS bands to produce a four-band pan-sharpened MS (PS-MS) image with 1-m resolution. The PS-MS imagery retained the 11-bit quantization of the original data.

Both the 4-m MS and 1-m PS-MS image datasets were classified using the traditional supervised maximum likelihood approach. A more detailed classification of the urban landscape is possible from the high-resolution IKONOS imagery compared to medium-resolution MS image data (e.g. Landsat). Accordingly, the identification of fine-scale urban features (residential houses, individual trees, etc.) in the image can be achieved. The urban land cover classes used in this initial study were *Road*, *Building*, *Grass*, *Tree*, *Bare Soil*, *Water*, and *Shadow*. The *Shadow* class is required to minimize





**Figure 2-1.** One-meter resolution panchromatic IKONOS image of Columbia, Missouri.

the problem of shaded pixels in the urban environment. Essentially, building shadows would often be incorrectly classified as *Water* if the *Shadow* class were not explicitly adopted. As there are a variety of spectral responses within each land cover class, multiple training sites must be used for each class. This results in a land cover classification map that contains multiple sub-classes for each land cover type. After the classification was performed, the sub-classes within each land cover type were merged

together to obtain a single class for each land cover type. An accuracy assessment of the resulting classification was performed using reference pixels that were independent of the pixels used to train the classifier. The reference pixel datasets were generated via photo interpretation of the 1-m PS-MS IKONOS imagery. Approximately 175 randomly distributed test site polygons were manually digitized in the imagery. The Columbia dataset had 9,410 training pixels and 80,895 reference pixels, and the Springfield dataset had 13,602 training pixels and 184,056 reference pixels. Although the training and test sites were generated from the PS-MS imagery, they were referenced via the map coordinates to the lower resolution MS imagery. The resulting number of training and test pixels in the MS imagery are reduced accordingly. The same training and reference pixel sets were used for all classification results presented in this chapter.

Supervised maximum likelihood classifications were produced for both the 4-m MS and 1-m PS-MS images from both study locations. The confusion matrix, the overall accuracy, and the Kappa coefficient of agreement were computed for each classification. The overall accuracy was computed by dividing the number of correctly classified reference pixels by the total number of reference pixels. The Kappa coefficient adjusts the overall accuracy value by subtracting the estimated contribution of chance agreement between classified pixels and reference pixels [Campbell, 1996]. For a review of statistical techniques used to assess the accuracy of classifications of remotely sensed imagery see Congalton *et al.* [1983], Hudson and Ramm [1987], and Congalton [1991]. The overall accuracies and Kappa coefficients are presented in Table 2-1. The overall accuracies for the Springfield image were higher than those corresponding to the Columbia image for both the 4-m MS and the 1-m PS-MS datasets. This is most likely

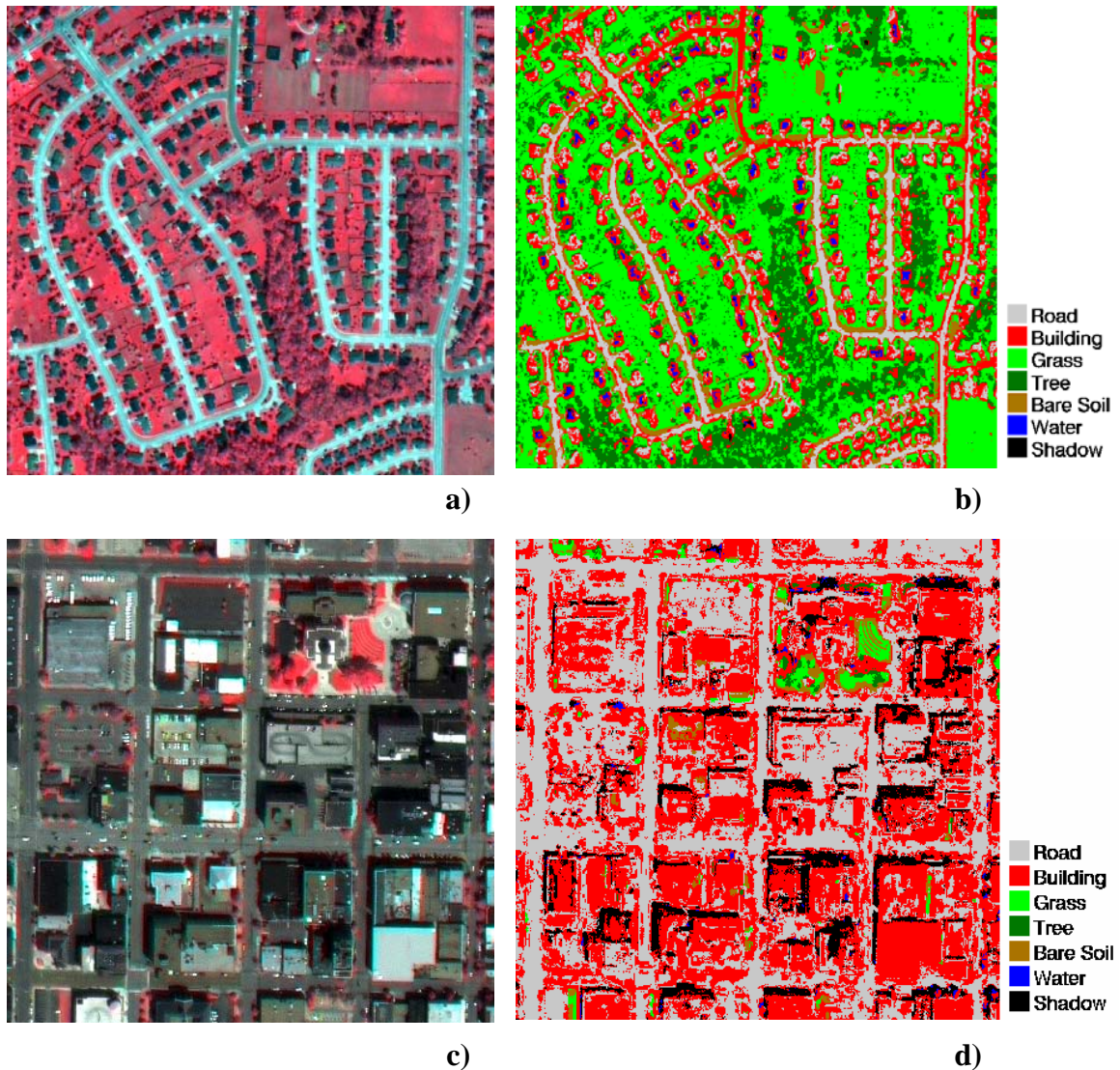
due to the presence of a small amount of haze in the Columbia image. The classification accuracies and Kappa coefficients of the 1-m PS-MS data are several percent higher than those of the 4-m MS data for both datasets, indicating that the higher spatial resolution pan-sharpened images produced by the color normalization method can be effectively used for classification purposes without loss of accuracy.

**Table 2-1**  
**Maximum Likelihood Classification Results for 4-m MS and 1-m PS-MS Image Datasets**

Pixel Size	Image	Overall Accuracy (%)	Kappa Coefficient
4-m	Columbia	78.6	0.731
1-m	Columbia	80.9	0.763
4-m	Springfield	85.7	0.818
1-m	Springfield	87.4	0.840

The confusion matrix for the PS-MS classification of the Columbia image is shown in Table 2-2. The largest source of error is due to misclassifications between the spectrally similar *Road* and *Building* classes, with 26% of the *Road* reference pixels classified as *Building* and 18% of the *Building* reference pixels classified as *Road*. The other major source of error is confusion between the *Grass* and *Tree* classes, with 16% of the *Grass* reference pixels classified as *Tree* and 11% of the *Tree* reference pixels classified as *Grass*. In addition, 26% of the *Water* reference pixels are classified as *Shadow*. Suburban and urban image subsets of the maximum likelihood classification for the Columbia dataset are shown in Fig. 2-2.

The confusion matrix for the PS-MS classification of the Springfield image shows similar misclassification characteristics. The confusion matrix for the Springfield PS-MS classification is shown in Table 2-3. As with the Columbia PS-MS classification, the



**Figure 2-2.** Maximum likelihood classification for (b) suburban area, (d) urban area from the Columbia, Missouri image subsets shown in (a) and (c), respectively. Note the significant misclassifications between the *Road* and *Building* land cover types.

largest source of error in the Springfield classification is caused by misclassifications between the *Road* and *Building* classes, with 30% of the *Road* reference pixels classified as *Building* and 31% of the *Building* reference pixels classified as *Road*. Unlike the classification of Columbia image, there is virtually no confusion between the *Grass* and *Tree* classes in the Springfield image. There is more spectral variation between these classes in the Springfield image because the image was acquired in the early fall time

period, resulting in less confusion between the classes. In addition, 24% of the *Water* reference pixels are classified as *Shadow*.

**Table 2-2**  
**Confusion Matrix for Maximum Likelihood Classification**  
**of 1-m PS-MS Columbia Image Dataset**

	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Bare Soil Ref.</b>	<b>Water Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	8109	2682	1	0	316	23	11,131	73
<b>Building</b>	3020	10,850	53	5	55	299	14,282	76
<b>Grass</b>	107	23	22,567	1225	0	0	23,922	94
<b>Tree</b>	54	57	4179	10,055	0	0	14,345	70
<b>Bare Soil</b>	177	49	10	0	8864	0	9100	97
<b>Water</b>	5	537	0	0	0	4961	5503	90
<b>Shadow</b>	15	712	0	0	0	1885	2612	
<b>Total</b>	11,487	14,910	26,810	11,285	9235	7168	80,895	
<b>%</b>	71	73	84	89	96	69		

**Table 2-3**  
**Confusion Matrix for Maximum Likelihood Classification**  
**of 1-m PS-MS Springfield Image Dataset**

	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Bare Soil Ref.</b>	<b>Water Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	21,219	10,667	57	20	33	0	31,996	66
<b>Building</b>	8968	24,165	82	4	1390	0	34,609	70
<b>Grass</b>	54	16	48,753	47	1	0	48,871	100
<b>Tree</b>	15	2	249	49,448	0	206	49,920	99
<b>Bare Soil</b>	0	5	0	0	13,266	0	13,271	100
<b>Water</b>	5	0	0	0	0	4009	4014	100
<b>Shadow</b>	9	0	0	0	0	1366	1375	
<b>Total</b>	30,270	34,855	49,141	49,519	14,690	5581	184,056	
<b>%</b>	70	69	99	100	90	72		

The *Road* and *Building* classes in both images and the *Grass* and *Tree* classes in the Columbia image are spectrally similar and have a significant amount of spectral overlap. This is the primary reason for the large number of misclassifications between these classes. Traditional supervised classification methods that only take into account spectral information, such as maximum likelihood, are unable to differentiate between these classes with a high degree of accuracy. Consequently, methods that utilize spatial

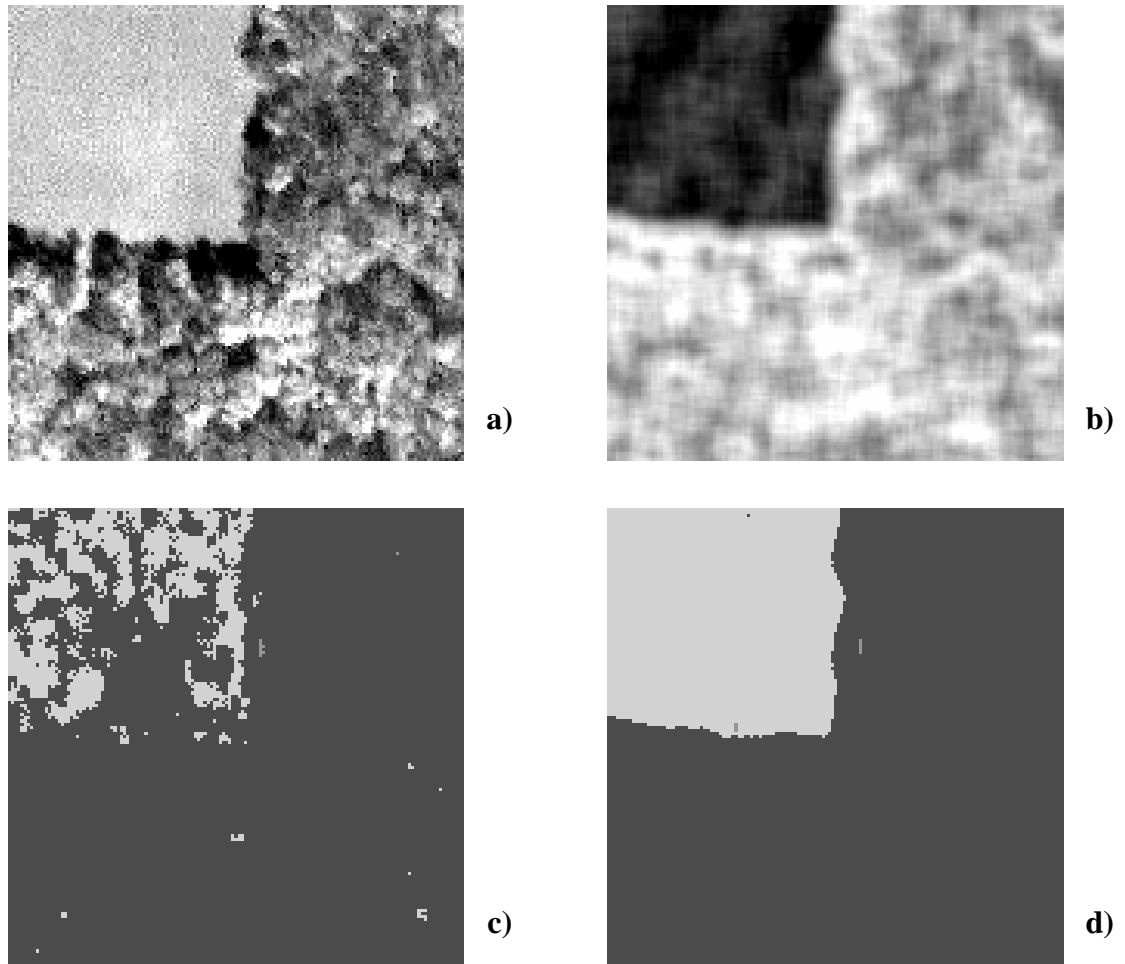
information in addition to spectral information are needed to produce more accurate land cover classifications of high-resolution image data over urban areas.

### **2.3 Spatial Feature Extraction**

By combining spatial information and spectral information, the amount of overlap between classes can be decreased, thereby yielding higher classification accuracies and more accurate urban land cover maps. For example, while the *Grass* and *Tree* classes can have similar spectral signatures, areas in the image covered with grass appear much more homogeneous than tree-covered areas. This difference in homogeneity between regions can be used to decrease the confusion between the classes. This is illustrated in Fig. 2-3 where an entropy texture measure is used to better differentiate between the *Grass* and *Tree* land cover types.

#### *A. Texture*

A variety of texture measures were evaluated for different window sizes to test their ability to improve the discrimination between the urban land cover classes. This was done by adding each texture image to the four PS-MS bands as an extra channel of data and then performing a maximum likelihood classification. The following first-order statistical texture measures were evaluated: entropy, data range, skewness, and variance. The texture features were calculated as defined in Eqs. 1-2 to 1-6. Each texture measure was calculated with a 5 x 5, 10 x 10, and 20 x 20 pixel window. These window sizes were chosen to be no larger than the objects of interest in the image from which the texture measures were designed to extract information. For that reason, a 20-meter wide



**Figure 2-3.** Effect of entropy texture measure on classification of *Grass* and *Tree* classes. (a) False color image subset. (b) 10 x 10 entropy texture measure. (c) Maximum likelihood classification of (a) (light gray = *Grass*, dark gray = *Tree*). (d) Maximum likelihood classification of PS-MS data + entropy.

window was the largest texture kernel size tested. While there were areas in the image, such as fields and large tree covered areas, that were much larger than this, the texture measures needed to be applicable to urban and suburban areas where many objects of interest are on the order of 10 – 20 meters in size. All of the texture measures discussed here were extracted from the PAN band of the IKONOS image datasets.

The average classification accuracies for the *Road* and *Building* classes and the *Grass* and *Tree* classes from the Columbia image are shown in Table 2-4. The first row

in the table is the average classification accuracies from the maximum likelihood classification of the PS-MS data with no added texture measures. The entropy texture measures using both a 10 x 10 and a 20 x 20 pixel window have a significant effect on the average classification accuracy of the *Grass* and *Tree* classes, where the classification accuracy of those classes increases approximately 10% in both cases. Although the classification accuracies of both the 10 x 10 and 20 x 20 entropy texture measures were essentially the same, the 10 x 10 window was chosen to help reduce edge effects associated with large texture windows [Ferro and Warner, 2002]. Several of the other texture measures show moderate increase in the accuracy of these classes, but not as large as the increase found when using the entropy texture measure. Most of the texture measures actually decrease the average classification accuracies for the *Road* and *Building* classes, and the best result (entropy 20 x 20) only yields a 1.5% increase over the PS-MS classification with no texture features.

### *B. 2-D Spatial Contextual Features*

It was found in Section 2.2 that the largest source of confusion in the classification of the high-resolution urban scenes is between the *Road* and *Building* classes. While many structures in the urban environment are spectrally similar such as roads, buildings and other impervious surfaces, roads are typically oriented along a long narrow ribbon, whereas other urban land cover types tend to have a more compact shape. Thus, by exploiting the differences in shape between these spectrally similar urban structures, it is possible to increase discrimination between these two classes. To provide a measure of shape, a feature extraction algorithm was developed to examine the context

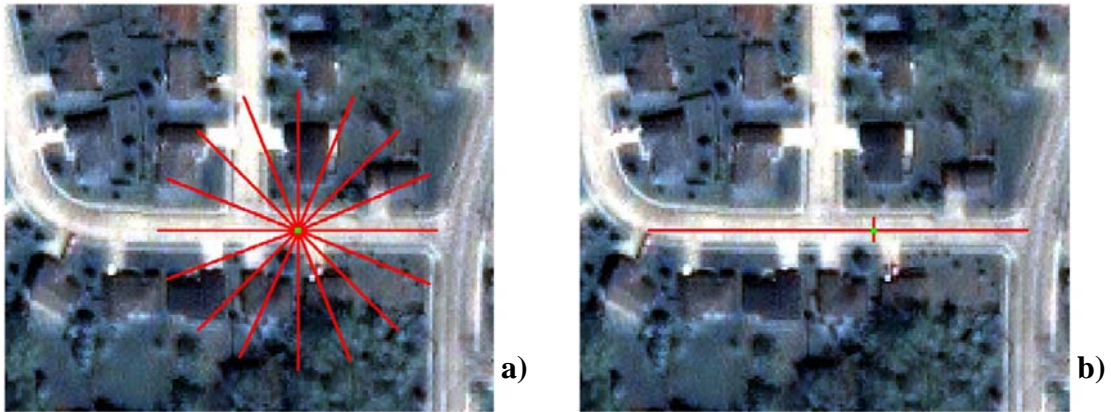


**Table 2-4**  
**Average Maximum Likelihood Classification Accuracies with Texture Information**  
**Included for 1-m PS-MS Columbia Image Dataset**

<b>Texture Feature</b>	<b>Window Size</b>	<b>Road – Building Avg. Accuracy (%)</b>	<b>Grass – Tree Avg. Accuracy (%)</b>
None	N/A	71.4	86.6
Entropy	5 x 5	70.4	87.8
Data Range	5 x 5	68.2	90.1
Skewness	5 x 5	70.7	90.7
Variance	5 x 5	65.7	91.4
Entropy	10 x 10	71.8	97.5
Data Range	10 x 10	69.1	89.9
Skewness	10 x 10	71.1	92.6
Variance	10 x 10	68.0	91.5
Entropy	20 x 20	72.9	97.3
Data Range	20 x 20	66.0	95.8
Skewness	20 x 20	69.9	84.3
Variance	20 x 20	70.2	87.8
2-D Contextual	N/A	76.2	77.9

around each pixel and compute quantitative estimates for the 2-D spatial extent (length and width) and directionality of groups of spectrally similar connected pixels. The algorithm is designed to extract this information from the PS-MS imagery. For each pixel in the image, the maximum and minimum length line segments passing through that pixel consisting of spectrally similar pixels are identified. These lengths are found by searching along  $N$  equally spaced straight-line segments passing through the central pixel from  $0^\circ$  to  $180^\circ$  with a step size of  $\Delta\theta$ . The Euclidean distance,  $d_e(\mathbf{x}^j, \mathbf{x}^n)$  is calculated (Eq. 1-7) between the spectrum of the central pixel,  $\mathbf{x}^j$ , and that of the other pixels,  $\mathbf{x}^n$ , along the various line segments. This calculation is performed for each of the  $2N$  line segments radiating from the central pixel, beginning with the pixel closest to the central pixel and moving outward until  $d_e$  exceeds a similarity threshold. Once all of the directions have been searched, the maximum and minimum length line segments are

identified and a number of parameters are stored for each pixel  $j$ . These are: (1) the maximum length value  $L^{max}_j$ , referred to as ‘length’; (2) the angle of the maximum length line segment  $\theta_j$ ; (3) the endpoints of the maximum length line segment  $EP^1_j$  and  $EP^2_j$ ; and (4) the minimum length value  $L^{min}_j$ , referred to as ‘width’. This process is illustrated in Fig. 2-4.



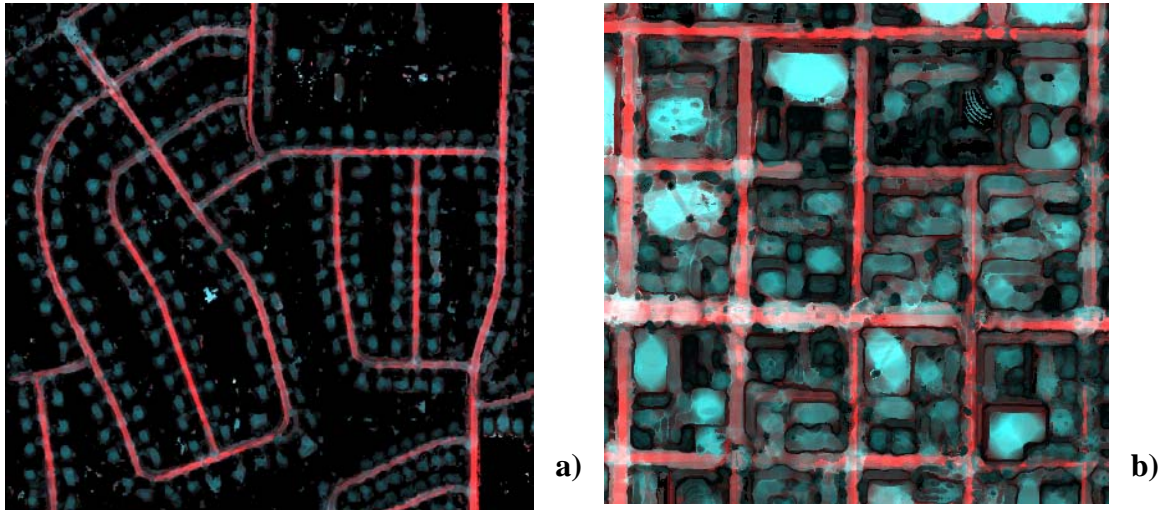
**Figure 2-4.** Illustration of 2-D spatial feature extraction process for the case of  $\Delta\theta=22.5^\circ$  ( $N=8$  search lines). The search directions are shown in (a) and the maximum and minimum length line segments are identified in (b).

Two parameters control the 2-D spatial feature extraction algorithm: the step size of the angle between each line segment to be searched,  $\Delta\theta$ , and the similarity threshold,  $d_{max}$ . The similarity threshold,  $d_{max}$ , largely determines the performance of the algorithm. If  $d_{max}$  is set too small or large, the length and width values will not accurately measure the spatial dimensions of objects in the imagery. A  $d_{max}$  value of 50 was found to work well for 11-bit IKONOS imagery. The azimuth sampling angle,  $\Delta\theta$ , must be small enough to capture the range of directions present in the road network, which is the most complex structure in the imagery. However, as  $\Delta\theta$  is decreased, the number of line segments being searched for each pixel increases, resulting in greater computational effort.  $\Delta\theta = 10^\circ$  was found to provide a dense enough azimuth sampling to characterize

all roads for the imagery examined in this study. It was found that if the image data were median filtered before the 2-D spatial features were extracted, the algorithm was able to more accurately measure the spatial dimensions of objects in the imagery. A 7 x 7 window for the median filter was found to work well. The kernel size for the median filter was chosen to be smaller than the desired objects being analyzed for contextual information (i.e. roads and buildings). However, the 7 x 7 window was large enough so that extremely fine-scale features in the image, such as automobiles and street linework, were removed.

The outputs of the 2-D spatial feature extraction algorithm applied to both an urban and a suburban scene are shown in Fig. 2-5. The length values are displayed in the red channel of an RGB display and width is displayed in the blue and the green channels. Vegetation pixels have been masked out so the effect of the contextual features on road and building pixels can be more clearly seen. Pixels that have large length values and small width values, such as road pixels, appear more red, while pixels with similar length and width values, such as building pixels, appear more blue. This algorithm was applied to the PS-MS Columbia image and the resulting length and width values were added as two bands of data to the four PS-MS bands and classified using maximum likelihood classification. These results are shown in the last row of Table 2-4. The average classification accuracy for the *Road* and *Building* classes increased by 5% when the 2-D contextual features were added. However, we note that the average classification accuracy for the *Grass* and *Tree* classes decreased by 9%. Finally, after inspection of the length and width probability density functions, it was found that they were not normally

distributed. Therefore, maximum likelihood classification is not the best choice for classification using this type of spatial feature.



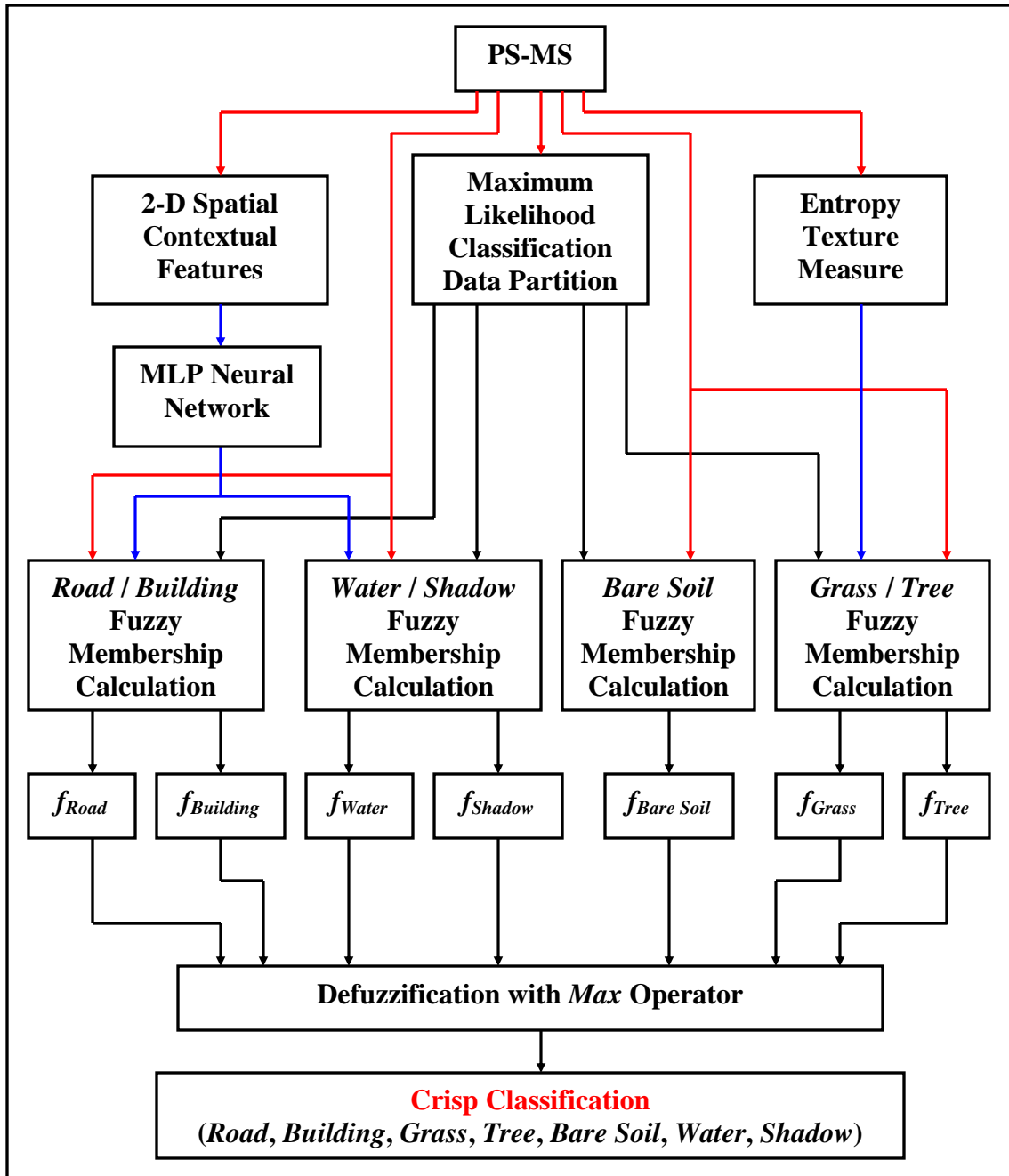
**Figure 2-5.** Length-width contextual measures of (a) suburban subset shown in Fig. 2-2a, and (b) urban subset shown in Fig. 2-2c.

## 2.4 Hierarchical Fuzzy Classification

As shown in the previous section, spatial measures extracted from the high-resolution imagery can help to decrease the number of misclassifications between the spectrally similar *Road / Building* and *Tree / Grass* classes. However, while one spatial feature might increase the classification accuracy between one set of classes, it might decrease the accuracy between another set using traditional classification methods. For example, the 2-D spatial contextual features discussed in the previous section increased the maximum likelihood classification accuracy between *Road* and *Building* by 5%, but the classification accuracy between *Grass* and *Tree* decreased by 9% (Table 2-4). The entropy texture measure increased the *Grass* and *Tree* maximum likelihood classification

accuracy by 10%, but this had almost no effect on *Road* and *Building* classification accuracy (Table 2-4).

Ideally, different classes should only be classified using the spatial measures best suited for those classes. Towards that end, we developed a fuzzy classification scheme that allows the image to be hierarchically classified using different spatial measures for different sets of classes. First, the maximum likelihood classification of the PS-MS data is used to split the data into four initial sets: *Grass-Tree*, *Road-Building*, *Water-Shadow*, and *Bare Soil*. A membership value for each class in each set is then calculated from membership functions generated from the PS-MS data plus the appropriate spatial measure. The 10 x 10 entropy texture measure is used for the *Grass-Tree* set, and the 2-D spatial contextual features are used for both the *Road-Building* and the *Water-Shadow* sets. As the classification accuracy of *Bare Soil* is already high, and no spatial measures were found to increase the classification accuracy of this class, only the PS-MS data is used to generate the membership value for the *Bare Soil* class. After membership values are calculated for each class in the set, the result is a fuzzy classification with each input pixel having a membership value in each class in the set. A crisp classification is generated in a defuzzification step using the *max* operator. A block diagram of this hierarchical fuzzy classification approach is shown in Fig. 2-6. The membership values for each class are calculated in parallel, so the pair classification order has no influence on the final outcome. This differs from a decision-tree approach [Theodoridis and Koutroumbas, 1998] where the pair-classification branching is done sequentially and the order of the pair branching is critical in the final classification outcome. Once divided into the initial sets, pixels can only be classified as one of the set members to which they



**Figure 2-6.** Block diagram of hierarchical fuzzy classification scheme. Red arrows indicate MS data, blue arrows indicate feature extraction data, and black arrows indicate processing steps.

belong. This does not have a negative impact on classifier performance as the sets are chosen to include only those classes that have the largest amount of spectral confusion.

As in the work by Melangi *et al.* [2000], the membership functions used for the PS-MS and entropy data are Gaussian shaped functions. The membership functions are defined using two parameters: the mean vector  $\mu^c$  and covariance matrix  $\Sigma_c$ , which are calculated from the training data (Eq. 1-7 and Eq. 1-17). The mean vector  $\mu^c$  is used to represent the ideal pixel in class  $C$ . If an input pixel  $\mathbf{x}$  has the value  $\mu^c$ , then it will have a membership value of 1.0, and as  $\mathbf{x}$  moves away from  $\mu^c$  the membership value decreases. The covariance matrix  $\Sigma_c$  governs the width of the function. The membership value in class  $C$  for the PS-MS and entropy data is calculated as

$$f_C'^{MS}(\bar{\mathbf{x}}) = \exp\left(-0.5 \cdot (\bar{\mathbf{x}} - \bar{\mu}^c)^T \cdot \Sigma_c^{-1} \cdot (\bar{\mathbf{x}} - \bar{\mu}^c)\right) \quad (2-1)$$

and this is a scalar value representing the degree to which input vector  $\mathbf{x}$  belongs to class  $C$ . In the case of the *Grass-Tree* set,  $\mathbf{x}$  is a five dimensional vector containing the PS-MS data and the 10 x 10 entropy texture measure. For the other three class sets (*Road-Building*, *Water-Shadow*, and *Bare Soil*) the input vector  $\mathbf{x}$  contains only the PS-MS data. If there are multiple training sites per class, as is the case here, the multiple membership values for each land cover class are simply combined using the standard fuzzy union operator (Eq. 1-20) so only the largest value of  $f_c'^{MS}(\mathbf{x})$  is retained for each land cover class. Once the membership value in each class has been calculated, a primitive fuzzy membership vector is formed for  $\mathbf{x}$

$$\bar{F}'_{MS}(\bar{\mathbf{x}}) = [f_1'^{MS}(\bar{\mathbf{x}}), f_2'^{MS}(\bar{\mathbf{x}}), \dots, f_N'^{MS}(\bar{\mathbf{x}})]^T \quad (2-2)$$

where  $N$  is the number of classes. After the membership values for the PS-MS and entropy data have been calculated for each class, they are rescaled to normalize the membership values and form the fuzzy membership vector

$$\vec{F}_{MS}(\vec{x}) = [f_1^{MS}(\vec{x}), f_2^{MS}(\vec{x}), \dots, f_N^{MS}(\vec{x})]^T \quad (2-3)$$

where

$$f_i^{MS}(\vec{x}) = \frac{f_i'^{MS}(\vec{x})}{\sum_{j=1}^N f_j'^{MS}(\vec{x})} \quad (2-4)$$

This normalization takes place within all classes. The vector  $F_{MS}(x)$  represents the degree to which  $x$  belongs to each land cover class in terms of the PS-MS and entropy data.

A second membership value is calculated for the pixels in the *Road-Building* and *Water-Shadow* sets using the 2-D spatial contextual features. The length and width values are not normally distributed, so Gaussian-shaped functions are not appropriate for the membership functions. Instead, the membership functions are learned using a multi-layer perceptron neural network with one hidden layer consisting of 10 neurons. The use of a neural network allows the membership functions to be learned from training data without any prior assumptions about the distribution of the data. The multi-layer perceptron was chosen for its ability to approximate arbitrarily shaped functions and because of its ease in training. The multi-layer perceptron is trained using the standard back-propagation algorithm [Haykin, 1999].

The membership functions for all the data could have been generated using the multi-layer perceptron. However this approach was not chosen because the spectral and entropy data were normally distributed and best represented with Gaussian-shaped



functions. After the neural network learns the membership functions from the training data of the 2-D spatial contextual features, membership values in the *Road-Building* classes and the *Water-Shadow* classes are found for the pixels in those partitions resulting in a fuzzy membership vector

$$\bar{F}_{LW} = [f_1^{LW}(\bar{x}), f_2^{LW}(\bar{x}), \dots, f_N^{LW}(\bar{x})]^T \quad (2-5)$$

where  $f_i^{LW}(\mathbf{x})$  is the membership value of pixel  $\mathbf{x}$  in the length-width membership function (generated by the neural net) for class  $i$ . The vector  $\mathbf{F}_{LW}(\mathbf{x})$  represents the degree to which  $\mathbf{x}$  belongs to each class in terms of the 2-D spatial contextual features. Because these contextual features contain no information useful for the characterization of the *Grass*, *Tree*, and *Bare Soil* classes,  $f_i^{LW}(\mathbf{x})$  is set to zero for those classes.

At this point each pixel has two fuzzy membership vectors,  $\mathbf{F}_{MS}(\mathbf{x})$  and  $\mathbf{F}_{LW}(\mathbf{x})$ . These two vectors are combined using the standard fuzzy union operator (Eq. 1-20) to produce a single fuzzy membership vector

$$\bar{F}(\mathbf{x}) = [f_1(\bar{x}), f_2(\bar{x}), \dots, f_N(\bar{x})] \quad (2-6)$$

where

$$f_i(\bar{x}) = \max \left\{ (1 - \alpha^{MS}) \cdot f_i^{MS}(\bar{x}), (1 - \alpha^{LW}) \cdot f_i^{LW}(\bar{x}) \right\} \quad (2-7)$$

and  $\alpha_i^{MS}$  and  $\alpha_i^{LW}$  are values between 0.0 and 1.0 representing the uncertainty in the PS-MS data and the 2-D spatial contextual features for class  $i$ , respectively. For the *Road-Building* and *Water-Shadow* sets, values of 0.35 and 0.1 were used for  $\alpha^{MS}$  and  $\alpha^{LW}$ , respectively. The values for  $\alpha^{MS}$  and  $\alpha^{LW}$  were initially set by examining the individual class accuracies achieved using the different spatial features (Table 2-4), and then adjusted through trial and error to achieve the best performance. For the *Grass-Tree* set and the *Bare Soil* class, values of 0.0 and 1.0 were used for  $\alpha^{MS}$  and  $\alpha^{LW}$ , respectively.

These uncertainty values reflect the fact that the 2D spatial feature measure does not improve the classification accuracies for these classes. Note however that  $f_i^{MS}(\bar{x})$  does include the texture measure (Eq. 2-1) to improve the discrimination between the *Grass* and *Tree* classes.

At this point, each input pixel  $x$  now has one membership value in each of the  $N$  classes. Since a crisp classification is desired, the fuzzy classification must be defuzzified to produce a single class label for each pixel in the image. Defuzzification is performed using the *max* operator such that  $x$  is classified as the class  $C$  with the highest membership value

$$Class = \arg \max_{j=1}^N \{f_1(\bar{x}), f_2(\bar{x}), \dots, f_N(\bar{x})\} \quad (2-8)$$

## 2.5 Fuzzy Pixel Classifier Results

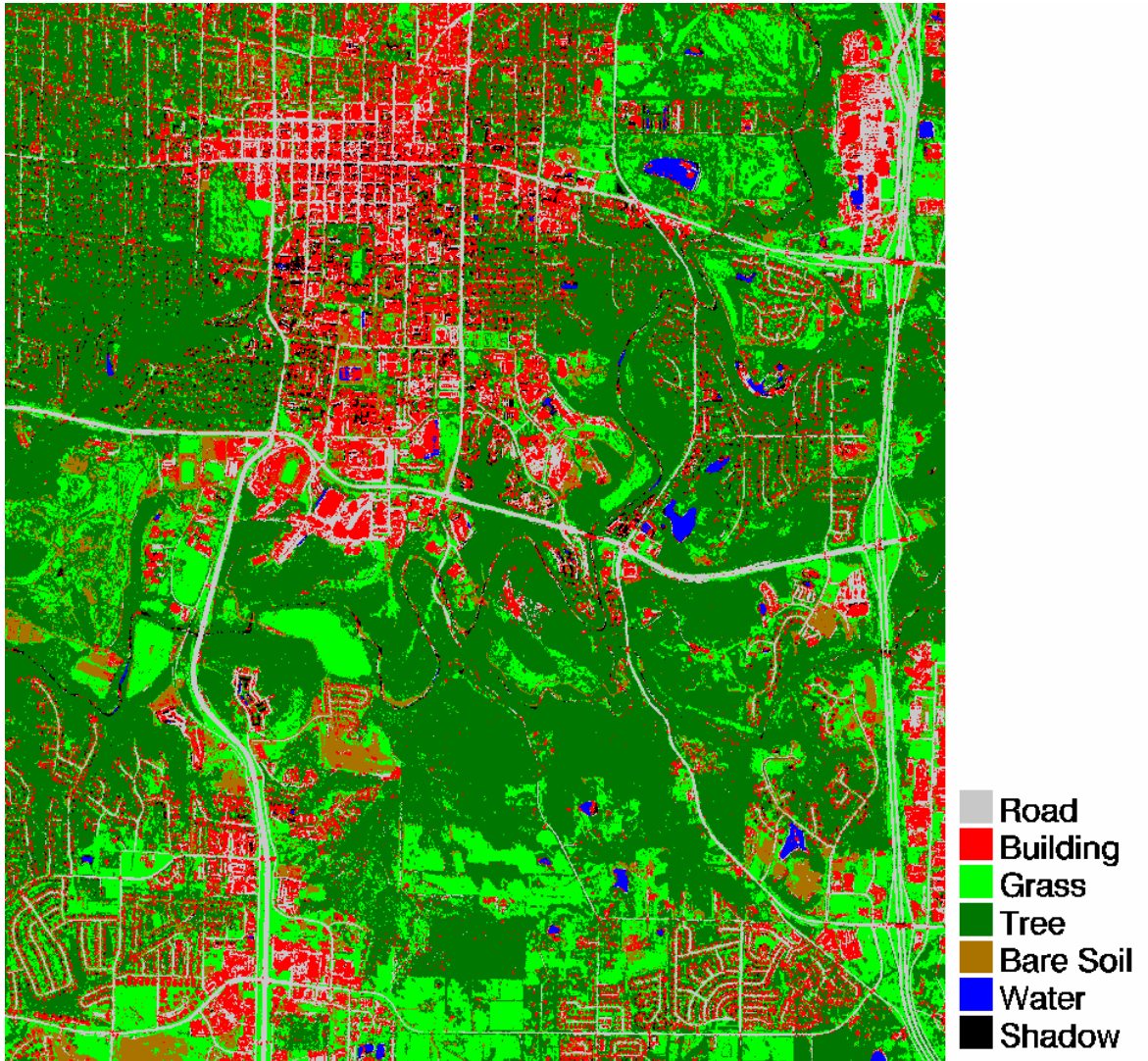
The hierarchical fuzzy classifier was applied to both the Columbia and Springfield image datasets using the same training data that generated the maximum likelihood classification results presented in Table 2-1. The land cover classification map of the Columbia imagery generated using the fuzzy classifier is shown in Fig. 2-7. The accuracy assessments of the crisp classifications using the hierarchical fuzzy classifier, generated using the same reference pixels used for the maximum likelihood classification accuracy assessments, are shown in Table 2-5. The overall accuracy of the Columbia image increased by approximately 11% over the maximum likelihood accuracy when the fuzzy classification scheme was implemented. Moreover, the Kappa coefficient increased by 0.146. The overall accuracy of the Springfield image increased by

approximately 8% over the maximum likelihood accuracy when the fuzzy classification scheme was implemented and the Kappa coefficient increased by 0.106.

**Table 2-5**  
**Overall Accuracies of Crisp Output of Fuzzy Classifier**

Image	Overall Accuracy (%)	Kappa Coefficient
Columbia	92.7	0.909
Springfield	95.8	0.946

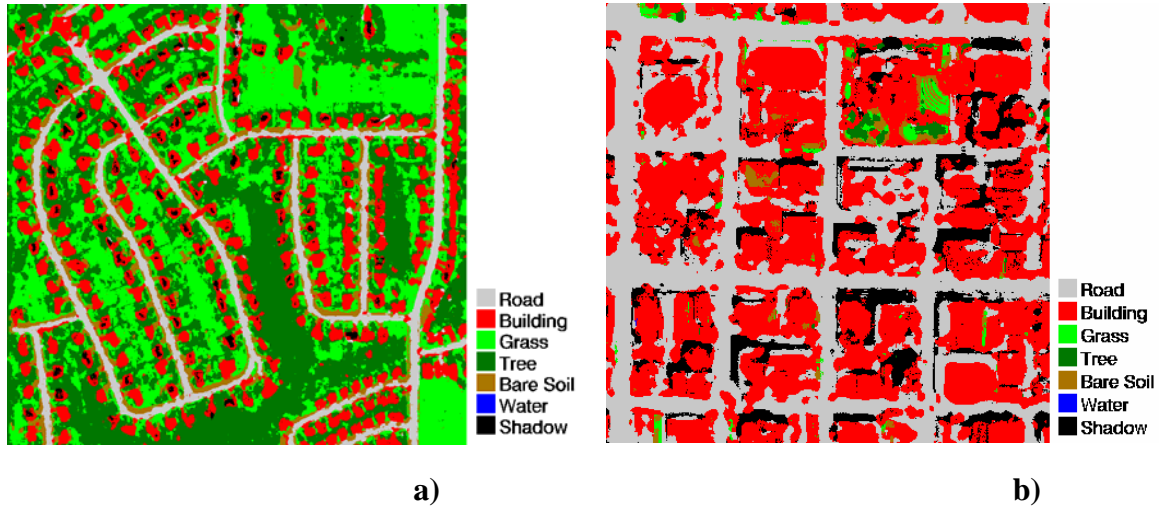
The confusion matrix for the crisp output of the fuzzy classification of the Columbia image is shown in Table 2-6. The average *Road-Building* classification accuracy increased from 71% to 86%, and the average *Grass-Tree* accuracy increased from 87% to 97%. In addition, the *Water* classification accuracy increased from 69% to 95%. Fig. 2-8 shows the crisp classification of suburban and urban area subsets from the Columbia image. As the classification maps show, the fuzzy classifier performs better in suburban areas than in urban areas, where the problems of spectral overlap and within class variance are most severe. However, when the classification maps in Fig. 2-8 and Fig. 2-2 are compared, it is clear that the fuzzy classifier outperforms the maximum likelihood classifier in both suburban and urban areas. The confusion matrix for the crisp output of the fuzzy classification of the Springfield image is shown in Table 2-7. The average *Road-Building* classification accuracy increased from 70% to 92%. The average *Grass-Tree* accuracy remained at 99%. In addition, the *Water* classification rate increased from 72% to 93%.



**Figure 2-7.** Crisp output of fuzzy classifier for Columbia, Missouri imagery shown in Fig. 2-1. Note the excellent delineation of road and building features.

**Table 2-6**  
**Confusion Matrix for Crisp Output of Fuzzy Classification**  
**of 1-m PS-MS Columbia Image Dataset**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	<i>Bare Soil</i> Ref.	<i>Water</i> Ref.	Total	%
<i>Road</i>	10,147	1019	2	1	231	16	11,416	89
<i>Building</i>	982	12,513	52	4	140	306	13,997	89
<i>Grass</i>	99	23	25,458	53	0	0	25,633	99
<i>Tree</i>	62	57	1288	11,227	0	0	12,634	89
<i>Bare Soil</i>	177	49	10	0	8864	0	9100	97
<i>Water</i>	2	318	0	0	0	6798	7118	96
<i>Shadow</i>	18	931	0	0	0	48	997	
<b>Total</b>	11,487	14,910	26,810	11,285	9235	7168	80,895	
<b>%</b>	88	84	95	99	96	95		



**Figure 2-8.** Crisp output of fuzzy classifier for (a) suburban scene, (b) urban scene from the Columbia, Missouri image subsets shown in Fig. 2-2a and Fig. 2-2c, respectively. Note the significant improvement over the maximum likelihood classification results also shown in Fig. 2-2.

**Table 2-7**  
**Confusion Matrix for Crisp Output of Fuzzy Classification**  
**of 1-m PS-MS Springfield Image Dataset**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	<i>Bare Soil</i> Ref.	<i>Water</i> Ref.	Total	%
<i>Road</i>	28,967	4215	12	1	423	0	33,618	86
<i>Building</i>	1220	30,617	127	23	1000	0	32,987	93
<i>Grass</i>	54	16	48,755	2	1	0	48,828	100
<i>Tree</i>	15	2	247	49,493	0	206	49,963	99
<i>Bare Soil</i>	0	5	0	0	13,266	0	13,271	100
<i>Water</i>	6	0	0	0	0	5183	5189	100
<i>Shadow</i>	8	0	0	0	0	192	200	
<b>Total</b>	30,270	34,855	49,141	49,519	14,690	5581	184,056	
<b>%</b>	96	88	99	100	90	93		

For comparison purposes, the hierarchical fuzzy classifier was also applied to the 4-m MS Columbia dataset. The same training and reference sites used for the Columbia PS-MS dataset were adopted, however the number of training and reference pixels decreased according to the decrease in image resolution. All feature extraction and classification parameters were kept the same. Also, the imagery was not smoothed with a median filter prior to application of the length-width feature extraction. The entropy texture measure was calculated using a 7 x 7 pixel window (28 x 28 m). This window size was the best compromise between minimizing edge effects and still extracting usable information from the objects of interest in the image.

The classification accuracies for the 4-m MS Columbia dataset are presented in Table 2-8. The confusion matrices from the maximum likelihood and fuzzy classifications of the 4-m MS Columbia dataset are shown in Tables 2-9 and 2-10 for comparison, respectively. As was the case with the 1-m PS-MS data, the hierarchical fuzzy classification accuracy for the 4-m MS data is higher (~6%) than the maximum likelihood classification accuracy. However, the increase in accuracy from the maximum likelihood classification to the fuzzy classification was larger (~12%) for the 1-m PS-MS

data. The confusion between the Road and Building classes is decreased, however there is little change in the classification accuracies of the *Grass* and *Tree* classes. The most likely explanation for this is that the texture information useful for discrimination between these two classes is contained primarily in the 1-m resolution PAN band. Thus, it is clear that the 1-m PS-MS imagery is better suited for urban land cover mapping than the 4-m MS imagery by itself.

**Table 2-8**  
**Classification Accuracies for 4-m MS Columbia Image Dataset**

	<b>Overall Accuracy (%)</b>	<b>Kappa Coefficient</b>
Fuzzy Classifier	83.5	0.791
Max. Likelihood	78.6	0.731

**Table 2-9**  
**Confusion Matrix for Maximum Likelihood  
Classification of 4-m MS Columbia Image Dataset**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	<i>Bare Soil</i> Ref.	<i>Water</i> Ref.	<b>Total</b>	<b>%</b>
<i>Road</i>	521	135	11	0	57	3	727	72
<i>Building</i>	183	765	23	46	36	18	1073	71
<i>Grass</i>	6	6	1560	273	0	0	1845	85
<i>Tree</i>	0	2	47	412	0	0	461	89
<i>Bare Soil</i>	9	4	0	0	487	0	500	97
<i>Water</i>	0	30	0	0	0	269	299	90
<i>Shadow</i>	0	29	0	0	0	173	202	
<b>Total</b>	719	971	1641	731	580	463	5105	
<b>%</b>	72	79	95	56	84	58		

**Table 2-10**  
**Confusion Matrix for Crisp Output of Fuzzy**  
**Classification of 4-m MS Columbia Image Dataset**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	<i>Bare Soil</i> Ref.	<i>Water</i> Ref.	Total	%
<i>Road</i>	559	79	0	16	39	3	696	80
<i>Building</i>	145	821	34	30	54	18	1102	74
<i>Grass</i>	6	6	1579	278	0	0	1869	84
<i>Tree</i>	0	2	28	407	0	0	437	93
<i>Bare Soil</i>	9	4	0	0	487	0	500	97
<i>Water</i>	0	47	0	0	0	409	456	90
<i>Shadow</i>	0	12	0	0	0	33	45	
<b>Total</b>	719	971	1641	731	580	463	5105	
<b>%</b>	78	85	96	56	84	88		

A majority filter was implemented to operate on the *Water*, *Shadow*, *Road* and *Building* classes to increase the accuracy of the fuzzy classification result and clean up the appearance of the classification image. A majority filter operates by extracting a window of pixels around the pixel of interest and reclassifies the central pixel as the class with the largest number of pixels in the window. The majority filter was first applied to the *Water* class, but instead of allowing the *Water* pixels to be reclassified as any class, the *Water* pixels were only allowed to be reclassified as *Water*, *Shadow*, *Road*, or *Building*. This was done to keep *Water* pixels from being reclassified into one of the vegetation classes.

After the *Water* pixels, the *Shadow* pixels were majority filtered next. The pixels were reclassified as *Water*, *Road*, or *Building* thus removing the *Shadow* class from the image. It is important to remove the *Shadow* class, as it is not a real urban land cover class. Finally, the *Road* and *Building* pixels were majority filtered and reclassified as *Road*, *Building*, *Water*, or *Bare Soil*. As was the case with the other majority-filtered classes, *Road* and *Building* pixels were not allowed to be reclassified as one of the



vegetation classes. The result of the majority filter post processing is a modest increase in classification accuracy of 1-2% and a more spatially coherent classification image.

## 2.6 Summary

The results presented in this chapter demonstrate the usefulness of high-resolution satellite imagery for urban land cover mapping and some of the shortcomings of conventional classification techniques such as maximum likelihood. It was found that maximum likelihood classification of high-resolution MS imagery over urban areas produced significant misclassification errors between spectrally similar classes such as *Road* and *Building* classes. Different spatial measures such as texture and contextual methods were investigated and found to increase the discrimination between certain spectrally similar classes. In particular, entropy first-order statistical texture measure (10 x 10 window) and the 2-D spatial contextual features were found to increase discrimination between the *Grass-Tree* and *Road-Building* classes, respectively.

Finally, a hierarchical fuzzy classification method was developed that utilized both the spectral and spatial information to classify the data. The fuzzy classification approach is based on a framework that allows for different groups of classes to be classified using the features best suited for discrimination between those classes. This alleviates the problem of features simultaneously decreasing the confusion between one set of classes and increasing it for another set. The classification accuracies of the fuzzy classifier were approximately 10% greater than the maximum likelihood classification results for 1-m PS-MS image datasets. Accordingly, there were significant decreases in the number of misclassifications between spectrally similar classes.

The classification techniques and results discussed in this chapter were presented at the *2002 International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2002] and were published in the *IEEE Transactions on Geoscience and Remote Sensing - Special Issue on Urban Satellite Remote Sensing* [Shackelford and Davis, 2003c]. The main contributions of this research to the body of literature on classification of high-resolution satellite imagery of urban areas is the introduction of the 2-D spatial contextual features for shape characterization and the fuzzy classification framework, where features are selectively applied only to those classes for which they are useful. The only other work on per-pixel land cover classification of high-resolution satellite imagery over urban areas [Benediktsson *et al.*, 2003] reported an overall classification accuracy of 95%, which compares well with the results presented here. However, many of the land cover classes used in the study were not true land covers (*Open Spaces, Wastelands, Open Areas*, etc.). Thus, our results while comparable in overall accuracy are more appropriate for true land cover mapping.

## Chapter 3

# Combined Pixel and Object-Based Fuzzy Urban Land Cover Classification

### 3.1 Introduction

Automated and semi-automated land cover classification can be accomplished using either pixel-based or object-based approaches. Pixel-based classification schemes seek to identify the class of each pixel in the imagery by comparing the  $n$ -dimensional data vector for each pixel with the prototype vector for each class. The data vectors typically consist of a pixel's gray-level values from multispectral (MS) channels and/or textural and contextual measures that have been computed from those channels. Object-based approaches do not operate directly on individual pixels but on objects consisting of many pixels that have been grouped together in a meaningful way by image segmentation. In addition to spectral and textural information for pixel-based classification methods, image objects also allow shape characteristics and neighborhood relationships to be used for the object's classification. However, the success of object-based classification approaches is very dependent on the quality of the image segmentation.

The fuzzy pixel-based classification approach presented in Chapter 2 produces a land cover classification map that is more accurate than the maximum likelihood

classifier by allowing different sets of classes to be classified using different spatial features in a hierarchical format. However, there are still significant classification errors in dense urban areas (see Fig. 2-8). In the urban land cover map generated from that technique, all man-made structures are classified as either *Road* or *Building*. This classification scheme is appropriate in suburban areas where residential homes and streets dominate the landscape. However, in dense urban areas there are significant amounts of non-road, non-building impervious surface land cover, such as parking lots, concrete plazas, and large sidewalks, and it is very desirable to differentiate between these surfaces, roads, and buildings. The fuzzy pixel-based classifier is unable to correctly identify these areas because they are spectrally similar to both the *Road* and *Building* classes, and the information extracted by the 2-D spatial contextual features is not detailed enough to discriminate between the *Building* and *Impervious Surface* classes. Consequently, even though the average classification accuracies reported for the fuzzy pixel-based classifier are around 86% for the *Building* class, this is not the true accuracy in dense urban areas as many parking lots, for example, are incorrectly classified as *Building*.

In this chapter, we adopt a segmentation and object-based classification approach to further refine the fuzzy pixel-based urban classification by adding an *Impervious Surface* class to identify non-road, non-building impervious surface. Because this land cover class is not found to a large extent in residential areas, and the fuzzy classifier performs well in those areas, the rest of this chapter will focus on the classification of dense urban areas only.

The classifier presented in this chapter is a combination of both the pixel-based and object-based approaches. Individual image pixels are first classified using the hierarchical fuzzy classifier introduced in Chapter 2. The imagery is then segmented and features are derived from the resulting image objects. The pixel class labels and object features are used together to classify the image objects using a fuzzy logic rule base. By using an object-based fuzzy logic approach in addition to the per-pixel fuzzy classification, additional information that was not available in the pixel-based classification can be identified and used to improve the overall urban land cover classification result.

### **3.2 Segmentation**

Before object-based classification can be performed, the imagery must first be segmented into image objects. Segmentation of remotely sensed images is a difficult problem due to mixed pixels, spectral similarity, and the textured appearance of many land-cover types. A variety of segmentation techniques have been applied to remote sensing imagery with varying degrees of success (see Section 1.3-B). The image segmentation algorithm used in this study follows the approach of Baatz and Schape [2000] where the segmentation is accomplished by region merging. The segmentation is initialized with each pixel in the image as a separate segment, and as the procedure progresses, segments are merged together. The decision to merge two segments together is based on the increase in heterogeneity of the new segment,  $h_{diff}$ , when compared to its constituent segments. A merge is only allowed when the increase in heterogeneity is less than a user set value,  $\varepsilon$ , called the scale parameter. As  $\varepsilon$  is increased, the size of segments

found in an image will increase. The procedure stops when there are no possible merges with a value of  $h_{diff}$  less than  $\varepsilon$ . The method used for finding two potential segments for a merge is local mutual best fitting. For a segment  $A$ , a neighboring segment  $B$  is found that has the smallest value of  $h_{diff}$  with  $A$ . For segment  $B$ , the neighboring segment  $C$  is found that has the smallest value of  $h_{diff}$  with  $B$ . If segment  $C$  and segment  $A$  are the same segment then segment  $B$  is the local mutual best fit for  $A$ . If not, then segment  $B$  replaces segment  $A$  and segment  $C$  replaces segment  $B$  and the procedure is repeated until the two local mutual best fitting segments are found. All segments in the image are examined in random order for each iteration to determine the local mutual best fit for each image segment. This procedure ensures even segment growth over the image.

Different measures of heterogeneity can be used that take into account both spectral and shape heterogeneity. Here, the spectral variance of each MS channel in the segment is used to measure spectral heterogeneity,  $h^b$ . The heterogeneity of channel  $b$  in segment  $j$ ,  $h_j^b$  is

$$h_j^b = \sigma_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{i,b}^j - \mu_b^j)^2 \quad (3-1)$$

where  $n_j$  is the number of pixels in segment  $j$  and  $\mu_j^b$  is the mean of band  $b$  in segment  $j$ . The shape heterogeneity,  $h^c$  is measured by comparing the length of the segment's perimeter,  $l_j$ , to the length of the perimeter of a square containing the same number of pixels. Thus,  $h^c$  quantifies the compactness of the segment. The shape heterogeneity of segment  $j$  is

$$h_j^c = \frac{l_j}{4 \cdot \sqrt{n_j}} \quad (3-2)$$

These heterogeneity measures favor segments with minimal spectral variance and compact shape. The increase in heterogeneity when two segments are merged is calculated as

$$h_{diff} = \sum_{k=1}^{m+1} w_k (n_1 (h_m^k - h_1^k) + n_2 (h_m^k - h_2^k)) \quad (3-3)$$

where  $h_m^k$  is the resultant heterogeneity of the merged segment for heterogeneity measure  $k$ ,  $h_1^k$  and  $h_2^k$  are the heterogeneities for the two segments being merged for measure  $k$ ,  $n_1$  and  $n_2$  are the number of pixels in each of the two segments being merged, and  $w_k$  is the weight for each heterogeneity measure. The weight determines the influence of each heterogeneity measure to the total increase in heterogeneity  $h_{diff}$ . For the specific case of IKONOS image data with four spectral channels,  $m = 4$  for the heterogeneity measures described above. The parameters  $h_m^k$ ,  $h_1^k$ ,  $h_2^k$ , and  $w_k$  for  $k = 1-4$  are the spectral heterogeneity measures and weights, and  $h_m^5$ ,  $h_1^5$ ,  $h_2^5$ , and  $w_5$  are the spatial heterogeneity measures and their weight.

The region-growing segmentation algorithm was applied to the IKONOS image and a subset of this showing the segmentation of a dense urban area (Fig. 3-1) is shown in Fig. 3-2. These results were obtained using a scale parameter  $\varepsilon = 700,000$  and weights  $w_k = 1$  for  $k = 1-4$  and  $w_5 = 400$ . Although the weight for the shape heterogeneity measure is much larger than the weights for the spectral heterogeneity measures, shape heterogeneity has less of an impact on  $h_{diff}$  than spectral heterogeneity because the increase in spectral heterogeneity when two segments are merged is much larger than the increase in shape heterogeneity. IKONOS imagery has very low contrast, and it was found that if the image is histogram equalized before segmentation, the segmentation results were greatly improved. While many of the road and impervious surface regions in the image are

broken into multiple segments, the segmentation successfully segments most of the buildings in the image as single segments. From this segmentation image it is then possible to use an object-based classification approach to differentiate between the *Building* and *Impervious Surface* classes.



**Figure 3-1.** One meter resolution PS-MS IKONOS image subset of dense urban area.



**Figure 3-2.** Segmentation of dense urban image shown in Fig. 3-1.

### 3.3 Fuzzy Logic Object Classification

Once a successfully segmented image is obtained, it is then possible to apply an object-based fuzzy logic classification to the segmentation image to assign a class label to each of the segments. Because we are dealing with image segments instead of pixels, object features such as shape and neighborhood information are available for use in the classification that were not available for the pixel-based fuzzy classifier. In addition, spectral statistics such as the mean and variance of the pixels in each segment can be calculated. Because the image has already been classified on a per-pixel basis, an initial



fuzzy classification of the image segments can be performed by analyzing the classes of the constituent pixels in each segment.

The goal of using an object-based approach at this stage in the processing is to improve the urban land cover classification in dense urban areas by discriminating between buildings and non-road impervious surface. This was not possible with the maximum likelihood or the fuzzy pixel-based approaches. The strategy employed for the discrimination between the *Building* and *Impervious Surface* classes is to label all pixels classified as *Building* from the fuzzy pixel-based classification as *Impervious Surface* and use subsequent object-based classification techniques to identify *Building* segments from within the *Impervious Surface* class. The object features used to identify *Building* segments are: (1) the classes of the constituent pixels that make up the segment, (2) morphological shape information, (3) location of potential *Building* segments with respect to *Shadow* segments, (4) mean panchromatic (PAN) brightness values of the segment, and (5) MS mean and variance values from the PS-MS data in each segment. Based on these five object features, a fuzzy logic rule base is used to discriminate between *Building* and *Impervious Surface* segments.

#### *A. Fuzzy Membership of Image Objects from Pixel Classes*

Initially, fuzzy membership values in each class are calculated for the segments based on the proportion of each class present in the segments from the per-pixel image classification. The class labels of all pixels classified as *Building* from the fuzzy pixel classification are changed to *Impervious Surface* so the segments initially have no membership in the *Building* class. Subsequent object-based processing is utilized to

identify *Building* segments from segments with a high membership value in the *Impervious Surface* class. An *S*-function [Klir, 1995] is used to calculate the membership values of a segment in each of the classes. First, the percentage,  $p_i$ , of each class present in the segment is calculated from the classification image, where  $i = 1, \dots, M$  and  $M$  is the number of classes. The membership value in each class,  $f_{c,i}(p_i)$ , is calculated from the *S*-function as,

$$f_{c,i}(p_i) = S(p_i; 0.0, 0.5, 1.0) \quad (3-4)$$

where

$$S(x; a, b, c) = \begin{cases} 0, & \text{for } x \leq a \\ \frac{1}{2} \left( \frac{x-a}{b-a} \right)^2, & \text{for } a < x \leq b \\ 1 - \frac{1}{2} \left( \frac{x-c}{c-b} \right)^2, & \text{for } b < x \leq c \\ 1, & \text{for } x > c \end{cases} \quad (3-5)$$

Segments consisting of pixels primarily from one class have membership values close to 1.0 in that class and membership values close to 0.0 in the other classes. Segments containing proportionately large numbers of pixels in several classes have no membership values close to 1.0 in any of the classes, reflecting the ambiguity of the information extracted from the constituent pixel classes in the segment. The subscript  $c$  in  $f_{c,i}(p_i)$  indicates that these membership values were calculated from the per-pixel classification of the image segments.

### B. Morphological Shape Processing

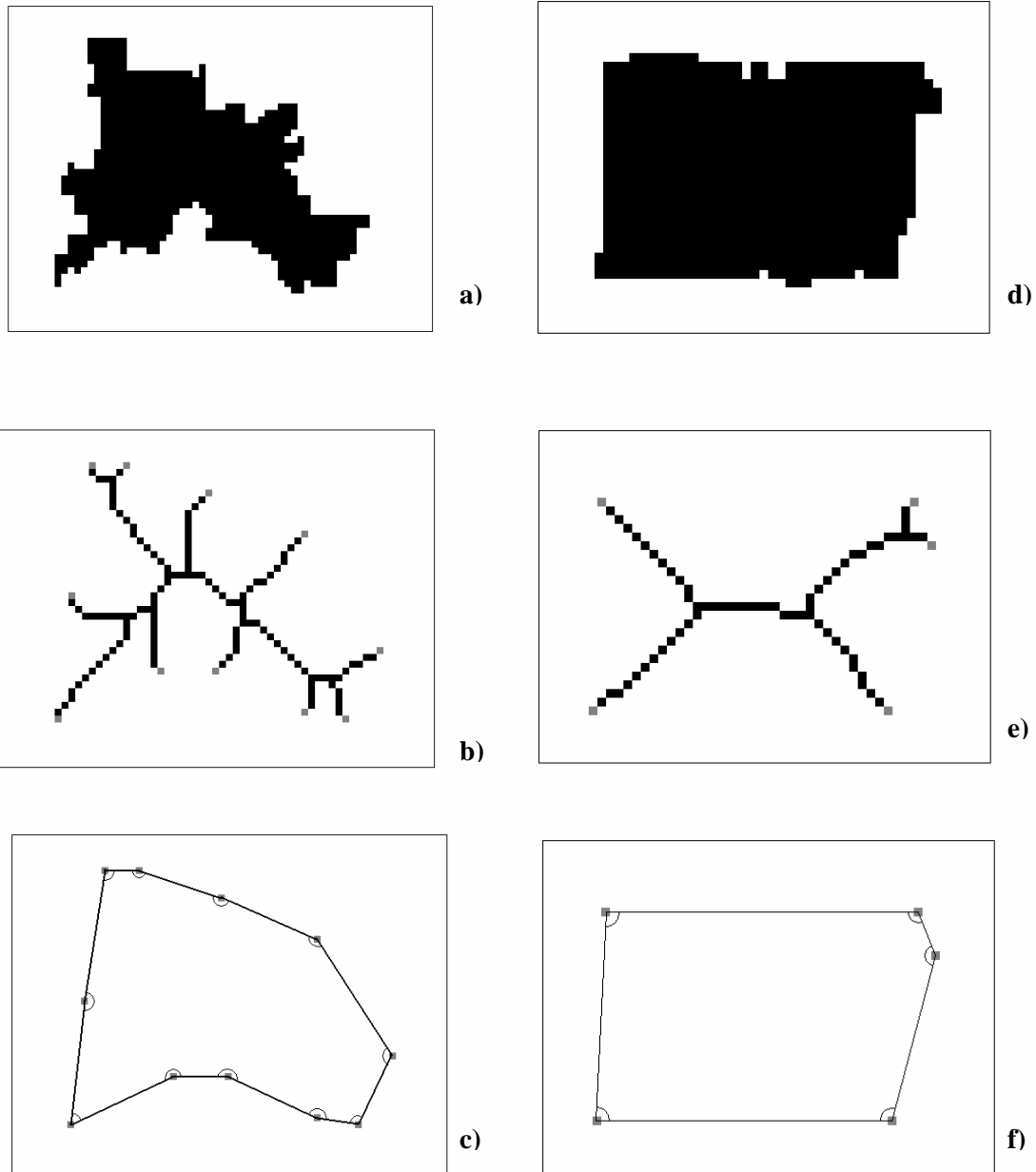
Shape information for each segment is measured using the morphological skeleton of the segment. Morphological skeletonization is a process for reducing a binary shape, in

this case a single segment, to a graph that largely preserves the extent and connectivity of the segment while discarding the foreground pixels in the segment. While a wide variety of skeletonizing algorithms exist, all do not however guarantee the production of a connected skeleton [Gonzalez and Woods, 2002]. The skeletonizing algorithm used here is based on a morphological thinning procedure and is guaranteed to produce a connected and one pixel thick skeleton [Jang and Chin, 1990]. The skeleton is found by cyclically applying morphological thinning operations to the segment until no further thinning is possible. The segment is iteratively thinned with eight structuring elements during each thinning cycle. The structuring elements used for the thinning are:

0	0	0
	1	
1	1	1

	0	0
1	1	0
	1	

and their 90° rotations. The 1's are the foreground and the 0's are the background elements of the structuring elements. During each thinning iteration, all eight structuring elements are individually passed over the segment as a convolutional mask, thereby removing the pixel at the center of the structuring element if the pixels of the segment exactly match the background and foreground elements of the structuring element. The thinning operations iterate until no further changes occur. This skeleton is very sensitive to small variations in the boundary of the original segment. In order to minimize the sensitivity to boundary variations, the original segments are initially smoothed by



**Figure 3-3.** Morphological skeleton processing for image segments shown in (a) and (d). Result of skeletonizing algorithm shown in (b) and (e) for image segments (a) and (d), respectively. The endpoints of the skeletons have been highlighted in gray. In (c) and (f), the skeleton endpoints have been connected by line segments and the endpoint angles identified.

performing a morphological closing operation using a 5-pixel diameter disk structuring element.

After the skeleton has been extracted from the segment, the endpoints of the skeleton are identified. Line segments are then found to connect the skeleton endpoints, and the angle between the two line segments connected to each endpoint is calculated. This process is illustrated in Fig. 3-3. The angles and line segment lengths are used as shape information to help classify the segment. Buildings are modeled as segments that are approximately rectangular in shape and also have a large membership value in the *Impervious Surface* class. An ‘approximately rectangular’ segment is modeled as a segment with the following attributes: about four endpoints with angles close to  $90^0$  and large separation, about two or less endpoints with angles much larger than  $90^0$ , and about two or less endpoints with angles much smaller than  $90^0$ . Because the attributes of an ‘approximately rectangular’ segment are imprecise and the shapes of buildings vary, fuzzy membership functions are used to measure how closely the endpoint angles and the line segment lengths between the endpoints match the different criteria.

First, three fuzzy membership values are calculated for each skeleton endpoint angle  $\theta_i$ : about  $90^0$ ,  $f_{\sim 90}(\theta_i)$ , larger than  $90^0$ ,  $f_{>90}(\theta_i)$ , and smaller than  $90^0$ ,  $f_{<90}(\theta_i)$ , where  $i = 1 \dots n_e$  and  $n_e$  is the number of endpoints in the segment skeleton. First,  $f_{\sim 90}(\theta_i)$  is calculated using a  $\pi$ -function [Klir, 1996] as

$$f_{\sim 90}(\theta_i) = \pi(\theta_i ; 70, 80, 85, 95, 100, 110) \quad (3-6)$$

where

$$\pi(x ; a, b, c, d, e, f) = \begin{cases} 0, & \text{for } x \leq a \text{ or } x \geq f \\ S(x ; a, b, c), & \text{for } a < x < c \\ 1, & \text{for } c \leq x \leq d \\ 1 - S(x ; d, e, f), & \text{for } d < x < f \end{cases} \quad (3-7)$$

Both  $f_{>90}(\theta_i)$  and  $f_{<90}(\theta_i)$  are calculated using  $S$ -functions as

$$f_{>90}(\theta_i) = S(\theta_i ; 95, 100, 110) \quad (3-8)$$

$$f_{<90}(\theta_i) = 1 - S(\theta_i ; 70, 80, 85) \quad (3-9)$$

The angle values controlling the shape of the membership functions in Eqs. 3-6, 3-8, and 3-9 were chosen so that angles matching the description of the membership functions would have membership close to 1.0, while angles deviating from the descriptions would have membership values approaching 0.0.

A membership value to quantify ‘large separation’ between the skeleton endpoints,  $f_{LS}(e_i)$ ,  $i=1, \dots, n_e$ , is calculated for each line segment connecting the segment skeleton endpoints as

$$f_{LS}(e_i) = S(\min[d_i, d_{i+1}]; 3, 5, 7) \quad (3-10)$$

where  $e_i$  is the  $i^{th}$  endpoint and  $d_i$  is the distance between endpoints  $e_{i-1}$  and  $e_i$ . ‘Large separation’ is defined in this case as being significantly larger than 5 meters. To identify endpoints that have both angles near  $90^0$  and large separation, the standard fuzzy intersection (Eq. 1-22) of  $f_{\sim 90}(\theta_i)$  and  $f_{LS}(e_i)$  is calculated using the *min* operator,

$$f_{\sim 90, LS}(e_i) = \min[f_{\sim 90}(\theta_i), f_{LS}(e_i)] \quad (3-11)$$

Fuzzy quantifiers [Klir, 1996] are used to calculate membership values for fuzzy sets describing ‘about four’ or ‘about two or less’. The membership value for a segment having about four endpoints with angles close to  $90^0$  and large separation,  $f_4$ , is calculated as

$$f_4 = \pi \left( \sum_{i=1}^{n_e} f_{\sim 90, LS}(e_i); 1.0, 1.5, 2.5, 5.5, 6.5, 7.0 \right) \quad (3-12)$$

The membership values for a segment having about two or less angles much larger than  $90^0$ ,  $f_{2>}$ , and having about two or less angles much smaller than  $90^0$ ,  $f_{2<}$ , are calculated as

$$f_{2>} = 1 - S\left(\sum_{i=1}^{n_e} f_{>90}(\theta_i); 2.0, 3.0, 4.5\right) \quad (3-13)$$

$$f_{2<} = 1 - S\left(\sum_{i=1}^{n_e} f_{<90}(\theta_i); 2.0, 3.0, 4.5\right) \quad (3-14)$$

Finally, the membership value for a segment being ‘approximately rectangular’,  $f_{rect}$ , is calculated using the standard fuzzy intersection of  $f_4$ ,  $f_{2>}$ , and  $f_{2<}$  with the *min* operator,

$$f_{rect} = \min[f_4, f_{2>}, f_{2<}] \quad (3-15)$$

### C. Segment Neighborhood Analysis

A common feature that buildings possess is that they cast shadows on the ground. The amount of shadows present in a high-resolution image will vary depending on the sun azimuth and elevation angles and the satellite sensor acquisition azimuth angle. As the sun elevation angle decreases from  $90^\circ$ , the length of the shadow buildings cast along the ground grows, and as the sun and sensor azimuth angles get further apart, the amount of shadows visible to the sensor will increase. If shadows are present in the image, they can be used to help identify *Building* segments. The IKONOS image used in this study had a sun elevation angle of  $61^\circ$ , a sun azimuth angle of  $139^\circ$ , and the sensor azimuth angle of  $352^\circ$ . As a result, building shadows are prominent as seen in the image subset shown in Fig. 3-1. Also note that the shadows are for the most part well segmented (Fig. 3-2), thereby allowing for successful neighborhood analysis.

The process for identifying segments as *Building* using neighborhood analysis of *Shadow* segments is as follows. First, segments with membership value of at least 0.5 in the *Shadow* class are identified. Then the segments bordering the potential *Shadow* segments in the direction of the sun azimuth angle are extracted. Segments with high

membership in the *Impervious Surface* class that share a border with these potential *Shadow* segment are likely *Building* segments. The search direction for segments neighboring potential *Shadow* segments must be quantized into one of eight directions because we are working on a digital image grid. Because of this, segments that are as much as  $45^\circ$  off the direction of the sun azimuth angle will be identified as potential *Building* segments. To avoid misclassifications, segments sharing a larger border along the direction of the sun azimuth angle with the potential *Shadow* segment are given higher membership values. The length of the border of each segment neighboring the *Shadow* segment in the direction of the sun azimuth angle is calculated and normalized by the largest border length. The membership value of the  $i^{th}$  segment,  $s_i$ , bordering a *Shadow* segment is calculated as

$$f_s(s_i) = S\left(\frac{l_i}{l_{max}}; 0.4, 0.6, 0.8\right), \quad (3-16)$$

where  $l_i$  is the length of the border shared with the *Shadow* segment and  $l_{max}$  is the length of the largest border shared with the *Shadow* segment. The  $f_s$  membership value is calculated for all segments bordering a segment in the direction of the sun azimuth angle with membership in the *Shadow* class of at least 0.5.

#### D. Spectral Object Features

While many buildings are spectrally similar to various impervious surface ground covers, some do have unique spectral signatures and can be identified from spectral features calculated from the constituent pixels within a segment. The spectral features calculated for the  $i^{th}$  band of the pixels in the segment are the mean,  $\mu_i$ , and the variance,



$\sigma_i^2$ . A number of the buildings in the image have very bright spectral responses, and a fuzzy membership function quantifying the brightness in the PAN band,  $f_{PAN}(\mu_p)$ , can be used to identify buildings of this type. Here  $\mu_p$  is the mean value of the PAN band of the pixels in the segment and,

$$f_{PAN}(\mu_p) = S(\mu_p ; 400, 550, 700) \quad (3-17)$$

Image segments with a large mean value in the PAN band will have membership close to 1.0 in  $f_{PAN}$ . The values governing the fuzzy membership function that identify a bright response in the PAN band were chosen by examining the spectral response of typical buildings that could be identified in this manner. Although these values were set manually, they could easily be calculated from training data.

Some buildings in the imagery, while they do not have a bright response in the PAN band, do have a unique spectral response in the MS bands, and this enables discrimination between the *Building* and *Impervious Surface* classes for these segments. Fuzzy membership functions for each MS band are defined using the mean,  $\mu_{t,i}$ , and variance,  $\sigma_{t,i}^2$ , calculated from *Building* training data, where  $i = 1, \dots, b$  and  $b$  is the number of MS bands. The membership value,  $f_{MS,i}$ , for the segment in each band is calculated using

$$f_{MS,i}(\mu_i) = \pi(\mu_i ; a, b, c, d, e, f) \quad (3-18)$$

where  $a = m_{t,i} - 0.8 \sigma_{t,i}$ ,  $b = m_{t,i} - 0.5 \sigma_{t,i}$ ,  $c = m_{t,i} - 0.3 \sigma_{t,i}$ ,  $d = m_{t,i} + 0.3 \sigma_{t,i}$ ,  $e = m_{t,i} + 0.5 \sigma_{t,i}$ , and  $f = m_{t,i} + 0.8 \sigma_{t,i}$ .

The parameters controlling the shape of the membership function were chosen such that segments with a mean value within one-half a standard deviation of the training data mean will have membership greater than 0.5. The membership values,  $f_{MS,i}$ , indicate

the degree of similarity between a segment and the *Building* training data for each MS band. These values are combined to form a single fuzzy membership for the segment,  $f_{MS}$ , by calculating the standard fuzzy intersection of the  $b$  membership values for the individual bands,

$$f_{MS} = \min[f_{MS,1}, \dots, f_{MS,b}] \quad (3-19)$$

Both  $f_{PAN}$  and  $f_{MS}$  are calculated for each image segment, indicating the degree of membership in the *Building* class in terms of the PAN and MS data, respectively.

#### E. Fuzzy Logic Classification of Object Features

The object membership values for the different *Building* attributes,  $f_{rect}$ ,  $f_S$ ,  $f_{PAN}$ ,  $f_{MS}$ , and  $f_{c,i}$ ,  $i = 1, \dots, M$ , are calculated for each segment in the image. Segments that do not border a potential *Shadow* segment have  $f_S$  set to 0.0. The identification of *Building* segments is accomplished with the use of a fuzzy logic rule base. The rule base used to calculate the membership value in the *Building* and *Impervious Surface* classes for the  $i^{th}$  segment in the image,  $s_i$ , is defined as follows:

**R1:** If  $s_i$  *Impervious Surface* membership is high AND  $s_i$  is approximately rectangular, then  $s_i$  membership in *Potential Building* is high.

**R2:** If  $s_i$  *Impervious Surface* membership is high AND  $s_i$  borders *Shadow* AND *Shadow* membership of bordering segment is high, then  $s_i$  membership in *Potential Building* is high.

**R3:** If  $s_i$  *Impervious Surface* membership is high AND  $s_i$  has bright PAN response, then  $s_i$  membership in class *Potential Building* is high.

**R4:** If  $s_i$  *Impervious Surface* membership is high AND  $s_i$  has *Building* MS response, then  $s_i$  membership in class *Potential Building* is high.

**R5:** If **R1** OR **R2** OR **R3** OR **R4** is high, then  $s_i$  membership in class *Building* is high.

**R6:** If  $s_i$  *Impervious Surface* membership is high AND  $s_i$  *Building* membership is NOT high, then  $s_i$  membership in *Impervious Surface* is high.

The standard fuzzy intersection and union operators are used for ‘AND’ and ‘OR’, respectively, and the standard fuzzy compliment is used for ‘NOT’. The fuzzy rule base is implemented as

$$f_{R,B}(s_i) = \min[f_{c,I.S.}(s_i), f_{rci}(s_i)] \quad (3-20)$$

$$f_{S,B}(s_i) = \min[f_{c,I.S.}(s_i), f_s(s_i), f_{c,S}(s_{sdw})] \quad (3-21)$$

$$f_{PAN,B}(s_i) = \min[f_{c,I.S.}(s_i), f_{PAN}(s_i)] \quad (3-22)$$

$$f_{MS,B}(s_i) = \min[f_{c,I.S.}(s_i), f_{MS}(s_i)] \quad (3-23)$$

$$f_{c,Building}(s_i) = \max[f_{R,B}(s_i), f_{S,B}(s_i), f_{PAN,B}(s_i), f_{MS,B}(s_i)] \quad (3-24)$$

$$f_{c,Imp. Surf.}(s_i) = \min[f_{c,I.S.}(s_i), 1 - f_{c,Building}(s_i)] , \quad (3-25)$$

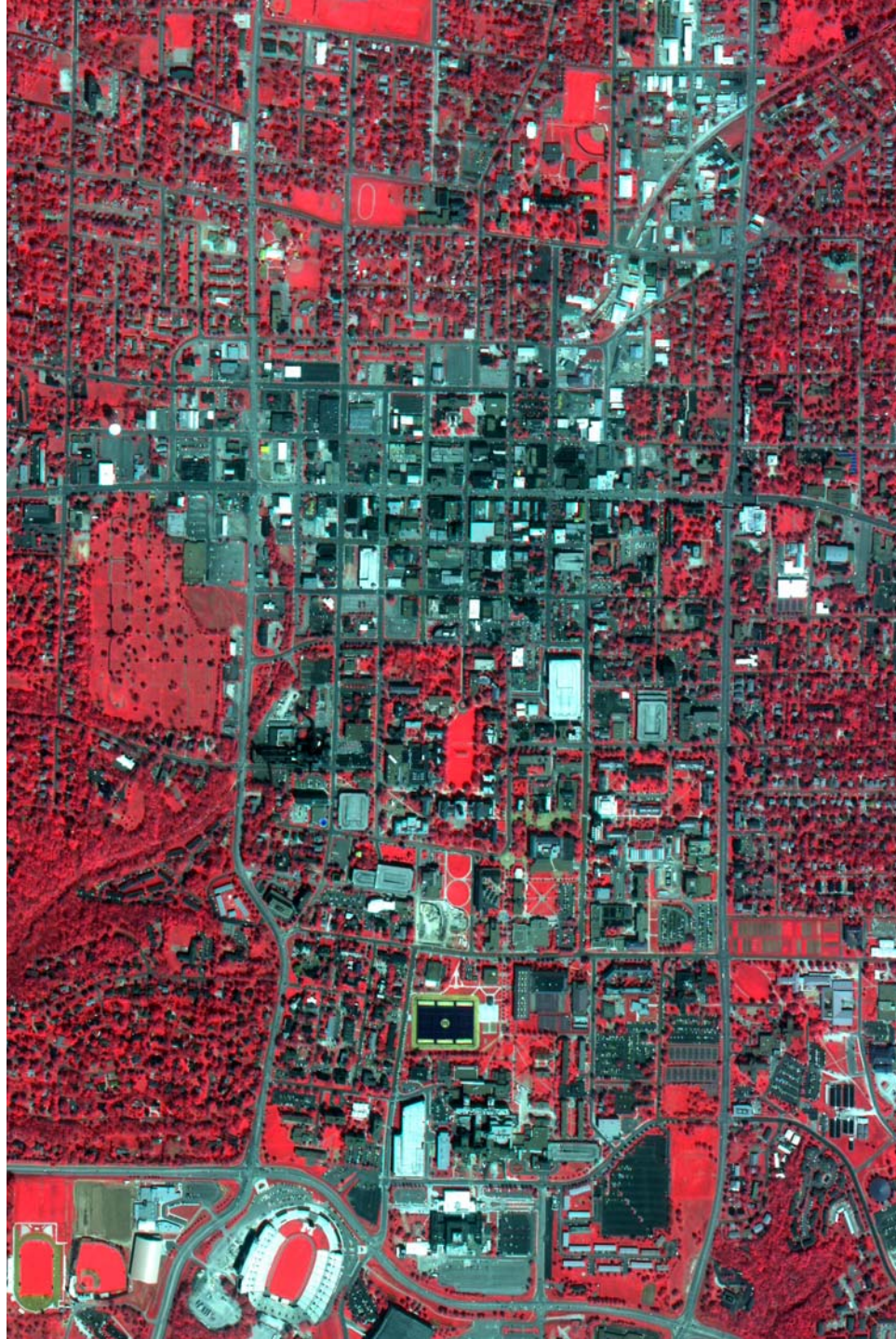
where  $f_{c,S}(s_{sdw})$  is the membership in the *Shadow* class of the segment bordering the segment being examined in the direction of the sun azimuth angle. The result of this processing is a fuzzy classification of the image segments, with each segment having membership values,  $f_{c,j}$ , in each of the eight classes: *Road*, *Building*, *Impervious Surface*, *Grass*, *Tree*, *Bare Soil*, *Water* and *Shadow*. Since a crisp classification is desired, the fuzzy classification must be defuzzified to produce a single class label for each image segment. Defuzzification is performed using the *max* operator such that segment  $s_i$  is classified as the class  $c$  with the highest membership value

$$Class(s_i) = \arg \max_{j=1}^M \{f_{c,1}(s_i), \dots, f_{c,M}(s_i)\} . \quad (3-26)$$

### 3.4. Fuzzy Object Classifier Results

The object-based fuzzy logic classification scheme was applied to a subset of the IKONOS image of Columbia, Missouri (Fig. 2-1). The test site is shown in Fig. 3-4 and consists primarily of dense urban land cover. The inputs to the classifier were the pixel-based fuzzy classification, and the PAN and PS-MS images. The prototype vectors used for training the *Building* MS fuzzy sets were the same used in training the pixel-based classifier (only those which fell within the test site were used). After the image was segmented, the object-based classification proceeded as described above. An accuracy assessment of the resulting classification was performed making use of reference pixels that were independent of the pixels used for training. The reference pixel datasets were generated via photo interpretation of the PS-MS IKONOS imagery. Because the test site is a subset of the image analyzed in Chapter 2, new reference pixels were identified. Approximately 130 randomly distributed test site polygons were manually digitized in the imagery. The dataset had 1,291 *Building* class training pixels and 51,368 reference pixels. Because this classification scheme was designed for the analysis of urban areas, the reference sites used for the accuracy assessment were all drawn from dense urban areas in the imagery. The test site had very few *Bare Soil* and no *Water* land cover, so no reference pixels were identified for those classes.

When each of the four object features was used separately to identify *Building* segments, only a fraction of the buildings were correctly identified. However, when all four were used in conjunction, 76% of the *Building* reference pixels were properly identified, while still correctly identifying 81% of the *Impervious Surface* reference pixels. Table 3-1 summarizes the effect of using different combinations of the object



**Figure 3-4.** One m resolution PS-MS IKONOS image of urban area test site.

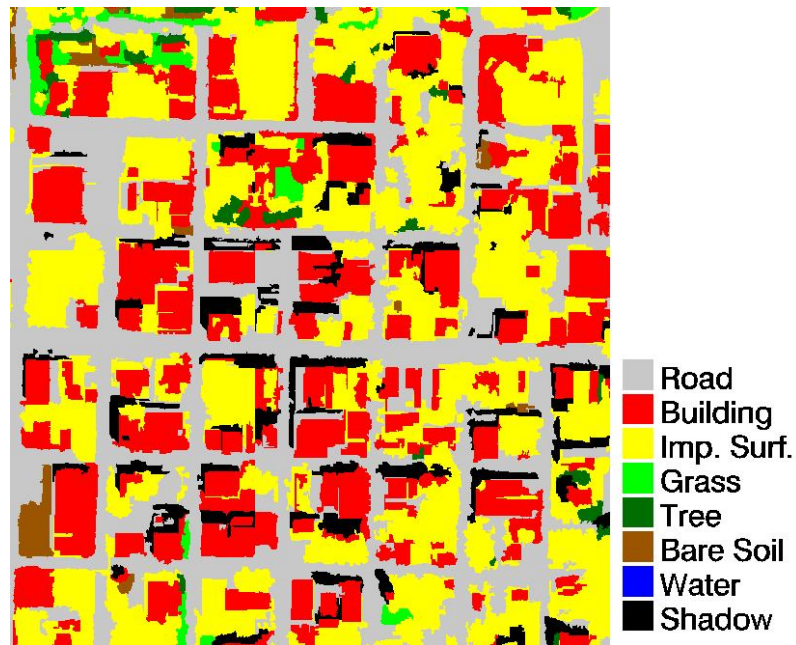
features on the accuracy of the *Building* and *Impervious Surface* classification. The single most effective object feature for discrimination between the *Building* and *Impervious Surface* classes is MS information, which correctly classifies 32% of the *Building* reference pixels. The *Building* class accuracy does not rise above 50% until at least three of the information sources are used. The results in Table 3-1 show that all four object-based features contribute to the identification of *Building* segments and each provides complimentary information. As the number of features is increased, the accuracy of the *Impervious Surface* class decreases. This occurs because the number of *Impervious Surface* pixels that are misclassified as *Building* increases as the number of correctly identified *Building* reference pixels increases. However, the average accuracy of the *Building* and *Impervious Surface* classes increases with the number of features, increasing the overall accuracy of the classification.

**Table 3-1**  
**Classification Accuracies for Different Combinations of Object-Based Features**

<b>Feature Combination</b>	<b><i>Building</i> Accuracy (%)</b>	<b><i>Impervious Surface</i> Accuracy (%)</b>	<b>Average Accuracy of <i>Building</i> and <i>Imp. Surf.</i> (%)</b>
Shape	24.9	87.4	56.2
Shadow	30.2	91.4	60.8
Pan	10.8	92.0	51.4
MS	32.0	84.4	58.2
MS + Pan	38.9	84.4	61.7
Shape + MS + Pan	54.9	81.0	68.0
Shadow + MS + Pan	64.5	83.8	74.2
Shape + Shadow + MS + Pan	76.1	81.0	78.6



The confusion matrix for the object-based classification of the urban area test site utilizing all four object-features is given in Table 3-2. The confusion matrix for the fuzzy pixel-based classification of the test site utilizing the reference pixels described in this section is given in Table 3-3. The *Road* class is very accurate with 99% of *Road* reference pixels correctly classified as *Road*. The *Road* class accuracy is slightly increased when using the object-based classifier as compared to the fuzzy pixel-based classifier. Because membership values of a segment calculated from the pixel-based classification are based on the proportional numbers of pixels of each class present in a segment, the segmentation operates as a majority filter if the segment is dominated by a single class. Misclassifications between the *Building* and *Impervious Surface* classes are



**Figure 3-5.** Object-based classification of dense urban area shown in Fig. 3-1 using segmentation image shown in Fig. 3-2 and fuzzy-logic pixel-based classification.

the largest source of error with 19% of *Building* reference pixels misclassified as *Impervious Surface*, and 11% of *Impervious Surface* reference pixels misclassified as *Building*. Comparing this to the pixel-based fuzzy classifier accuracies, the *Building* class accuracy appears to have decreased, dropping from 92% to 76%. It should be noted however that for dense urban areas, like the one shown in Fig. 3-1, the *Building* class accuracy reported for the pixel-based fuzzy classifier is an overestimate as there is no *Impervious Surface* class included in that classification scheme and non-road, non-building impervious surface makes up a significant portion of the landscape in dense urban areas. The object-based fuzzy classifier produces a more detailed urban land cover classification by including both a *Building* class and an *Impervious Surface* class. The classification image of the dense urban land cover subset image (Fig. 3-1) generated using this technique is shown in Fig. 3-5. The land cover classification map of the entire urban test site (Fig. 3-4) is shown in Fig 3-6. For comparison purposes, the hierarchical fuzzy pixel-based classification (Fig. 2-7) of the urban test site is shown in Fig 3-7.

**Table 3-2**  
**Confusion Matrix for Object-Based Classification of Urban Area Test Site**

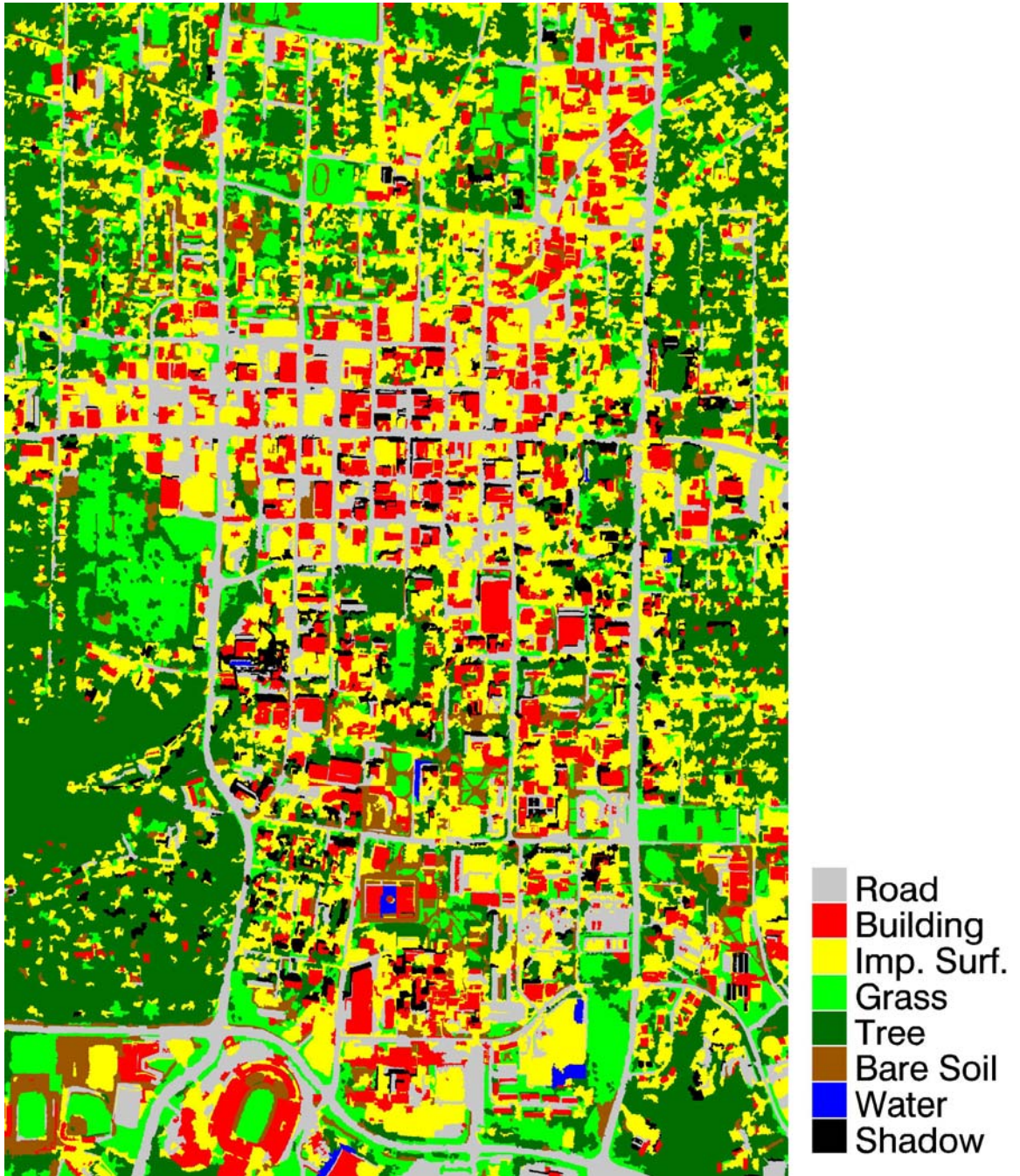
	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Imp. Surf.</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	Total	%
<i>Road</i>	10164	499	826	0	0	11489	88.5
<i>Building</i>	0	7845	1137	0	0	8982	87.3
<i>Imp. Surf.</i>	86	1959	8343	0	13	10401	80.2
<i>Grass</i>	0	0	0	9330	0	9330	100.0
<i>Tree</i>	0	0	0	887	10279	11166	92.6
<b>Total</b>	10250	10303	10306	10217	10292	51368	
<b>%</b>	99.2	76.1	81.0	91.3	99.9		89.5



**Table 3-3**  
**Confusion Matrix for Pixel-Based Classification of Urban Area Test Site**

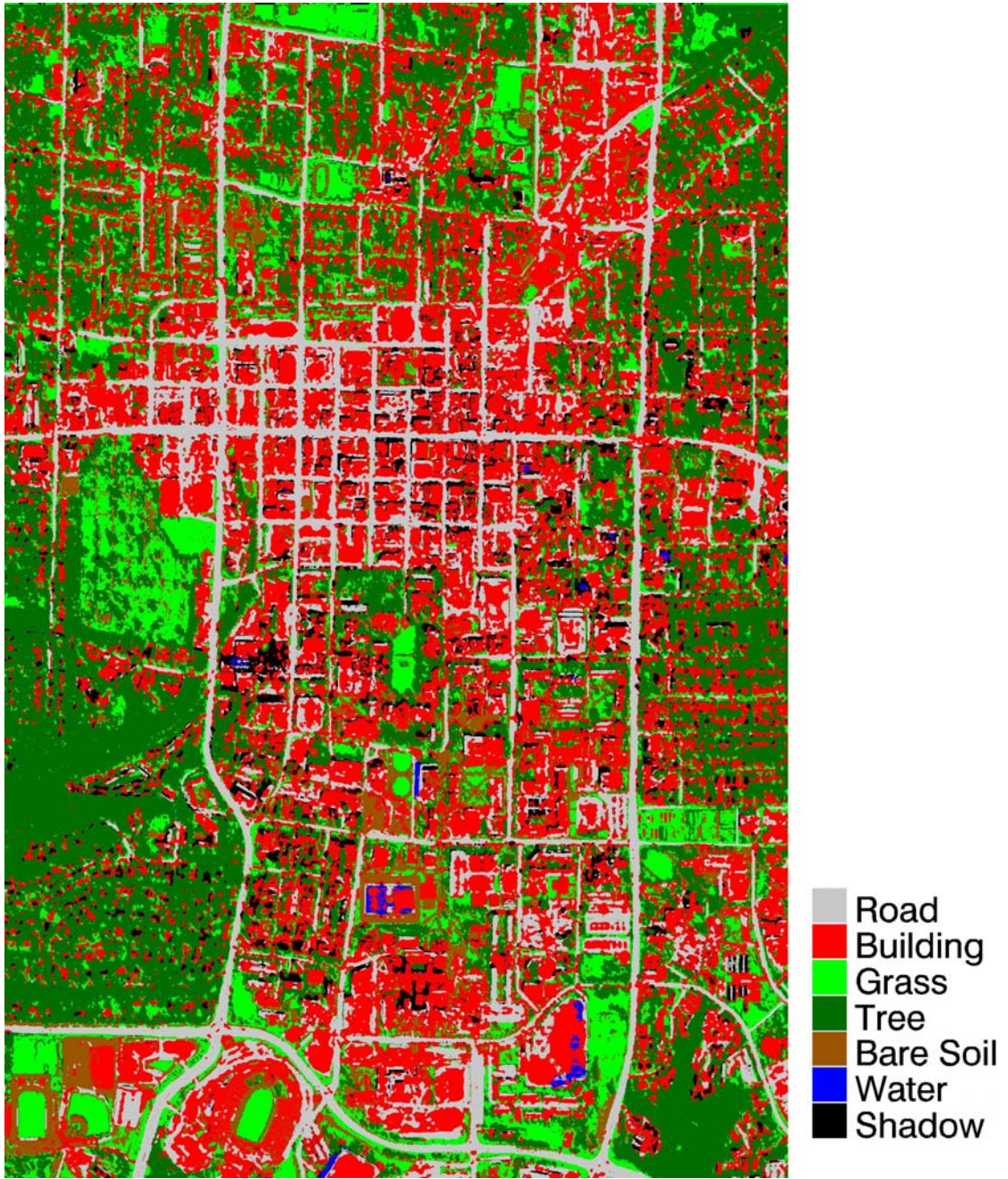
	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	Total	%
<i>Road</i>	10029	808	0	31	10868	92.3
<i>Building</i>	194	9309	0	248	9751	95.5
<i>Grass</i>	19	1	9540	112	9672	98.6
<i>Tree</i>	0	0	561	9901	10462	94.6
Total	10242	10118	10101	10292	40753	
%	97.9	92.0	94.5	96.2		95.2

Due to the methodology used to identify the *Building* class, buildings in the suburban areas along the edges of the test site are labeled as *Impervious Surface*. This object-based classification scheme was developed to discriminate between buildings and non-road impervious surface in areas of dense urban land cover, and accordingly is not suitable for classification of areas with suburban land cover. To address this issue, the *Building* class was subdivided into *Urban Building* and *Suburban Building*. The *Suburban Building* class is identified as image objects labeled as *Impervious Surface* with size less than 450 pixels and with vegetation comprising at least 25% of the land cover surrounding the segment. The percentage of vegetation surrounding the segment is determined by examining the class labels of the pixels within 10 m of the segment. The identification of the *Suburban Building* class is performed before the object features are examined for the identification of the *Urban Building* class, as described in Section 3.3 above. The land cover classification map of the urban test site (Fig. 3-4) with discrimination between the two building classes is shown in Fig 3-8. The results of applying this classification technique to an area with suburban land cover (Fig 3-9) are shown in Fig. 3-10.



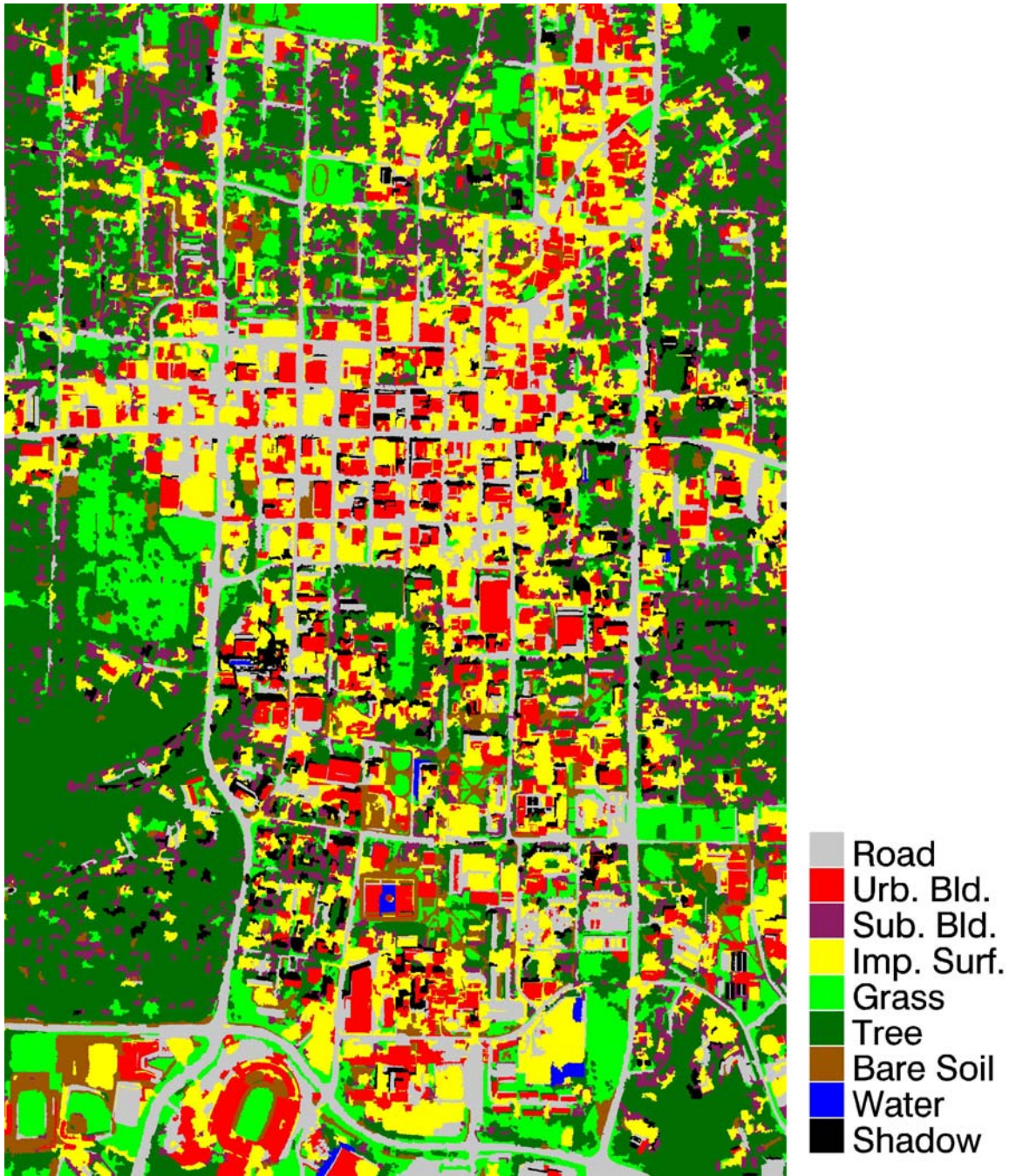
**Figure 3-6.** Fuzzy logic object-based classification of urban area shown in Fig. 3-4.





**Figure 3-7.** Hierarchical pixel-based fuzzy classification of urban area shown in Fig. 3-4.





**Figure 3-8.** Fuzzy logic object-based classification of urban area shown in Fig. 3-4 with differentiation between urban and suburban buildings.





**Figure 3-9.** One m resolution PS-MS IKONOS image of suburban area.



**Figure 3-10.** Fuzzy logic object-based classification of suburban area shown in Fig. 3-9 with differentiation between urban and suburban buildings.

### 3.5. Summary

A combined fuzzy pixel-based and object-based approach for urban land cover classification of high-resolution MS image data is presented in this chapter. The hierarchical fuzzy classifier presented in Chapter 2 classified the image on a per-pixel basis. Although this technique proved to be significantly more accurate than maximum likelihood classification, the pixel-based classification approach was only able to

successfully distinguish between the *Road* and *Building* classes in the urban environment. However, these classes are not sufficient to represent all land cover types present in dense urban areas. Accordingly, a non-road, non-building *Impervious Surface* class is also needed to represent features such as parking lots, concrete plazas, etc. To distinguish between these urban land cover classes, an object-based classification approach is introduced to examine features such as object shape and neighborhood and then classify the image objects using a fuzzy logic rule base.

To facilitate object classification, the imagery is first segmented with a region merging segmentation technique based on minimizing the increase in spectral and shape heterogeneity for each merge. The measure of spectral heterogeneity is the spectral variance of the object, and the compactness of the object quantifies the shape heterogeneity. This segmentation technique is able to correctly delineate buildings without over or under segmenting them. Several features are extracted from the image objects and used by the object-based classifier along with the fuzzy pixel-based classification. These features are the class labels of each segment's constituent pixels, shape information from the image objects, neighborhood analysis, and spectral statistics of the object. A shape model for the *Building* class, based on the skeleton of the image objects, is constructed using fuzzy membership functions, and the neighborhood analysis consists of examining the relationship between *Building* and *Shadow* segments. The image objects are classified with a fuzzy logic rule base. The fuzzy logic rule base classification approach is flexible in that more object features and rules can be easily added to the rule base. This classification approach produces an urban land cover map with a 76% accurate *Building* class and an 81% accurate *Impervious Surface* class.

The material presented in this chapter was published in the *IEEE Transactions on Geoscience and Remote Sensing* [Shackelford and Davis, 2003d]. The major contributions of this research are the combination of both pixel-based and object-based approaches into a single classifier, discrimination between very specific urban land cover classes (*Road, Building, Impervious Surface*), and the use of fuzzy sets to quantify shape information of image segments through the use of morphological skeletonization. While there are very few other works in the literature on object-based urban land cover classification, the overall accuracy of 90% obtained here (Table 3-2) is significantly better than the work by Kressler *et al.* [2001], where an overall accuracy of 71.5% was reported. In that study, the software package eCognition was used to perform segmentation and classification of a PS-MS IKONOS image. Ten detailed urban land cover classes were identified (*Water, Trees, Grassland, Fields, Sand, Buildings, Asphalt, Gravel, Gravel of Railtracks, Hardcourts*). Although buildings and several different types of road surfaces were classified, the classification scheme did not discriminate between roads and non-road impervious surface. The object-based classifier presented in this chapter is able to discriminate between these classes, producing a very detailed urban land cover map.

## **Chapter 4**

### **Automated Road Network Extraction**

#### **4.1 Introduction**

In the previous two chapters, techniques for classification of high-resolution imagery from urban areas were presented. A pixel-based hierarchical fuzzy classifier was presented in Chapter 2, and an object-based fuzzy logic classifier was presented in Chapter 3. Both of these techniques were shown to produce accurate urban land cover maps, and to be significantly more accurate than traditional maximum likelihood classification. In this chapter, the problem of automated road network extraction from high-resolution imagery of urban areas is addressed. As was discussed in Chapter 1, there are many applications for these features, such as urban growth management, transportation planning, emergency response, and automated navigation. Currently, the extraction of road networks from high-resolution imagery is done mostly by manual digitization, and this is both time consuming and expensive. Thus, automated and/or semi-automated methods for identification and extraction of road networks are therefore of great interest.

Automated extraction of roads from high-resolution imagery is a difficult task because of the complexity and the spatial and spectral variability of the road network. Roads can exhibit a variety of spectral responses due to differences in age and/or material and vary widely in physical dimensions. In addition, road networks in dense urban areas



typically have different geometric characteristics than those in suburban and rural areas. There are at least two general approaches for the automated/semi-automated extraction of road networks. Roads can be extracted directly from the digital imagery, or the road network can be extracted from a crisp or fuzzy classification of the data. Both approaches are examined in this chapter. Several techniques for extraction of the road network from a classification image are presented, followed by a fully automated method which extracts the road network directly from the digital imagery. All of the road network extraction algorithms presented here seek to extract a single pixel wide curvilinear response for each road estimating the location of the road centerline. The estimated centerline can be used for initialization of road edge detection techniques, or input directly into a GIS database as a vector line layer.

## **4.2 Accuracy Assessment**

Because the outputs of a pixel land cover classification and a vector road network extraction are fundamentally different, the confusion matrix technique used to assess the accuracy of the classification results presented in the previous chapters is not applicable here. To assess the road extraction accuracy, the length of the extracted road network that falls within a predefined distance of the centerlines of the reference road network is used for the calculation of several accuracy measures. The centerlines of the roads in the test sites are manually digitized to form the reference road network. This is more complete than the reference data used to test the classification results in Chapters 2 and 3, where randomly distributed reference polygons, independent of the training data, were selected in the image. The reference road network used to evaluate the road extraction

covers nearly all roads present in the image, providing a very accurate measure of the road extraction quality. A buffer is set up around the reference road centerlines, and the length of the extracted road network that falls within the buffer,  $L_{CE}$ , is determined. The length of the correctly extracted reference road network,  $L_{CR}$ , is determined in the same manner. The radius of the buffer,  $d_{buff}$ , is chosen to match the typical width of roads in the test site. Three measures are used to evaluate the accuracy of the extracted road network [Heipke *et al.* 1997], and these measures are defined as follows:

$$Completeness = \frac{L_{CR}}{L_R} \quad (4-1)$$

$$Correctness = \frac{L_{CE}}{L_E} \quad (4-2)$$

$$Quality = \frac{L_{CE}}{L_E + L_{UR}} \quad (4-3)$$

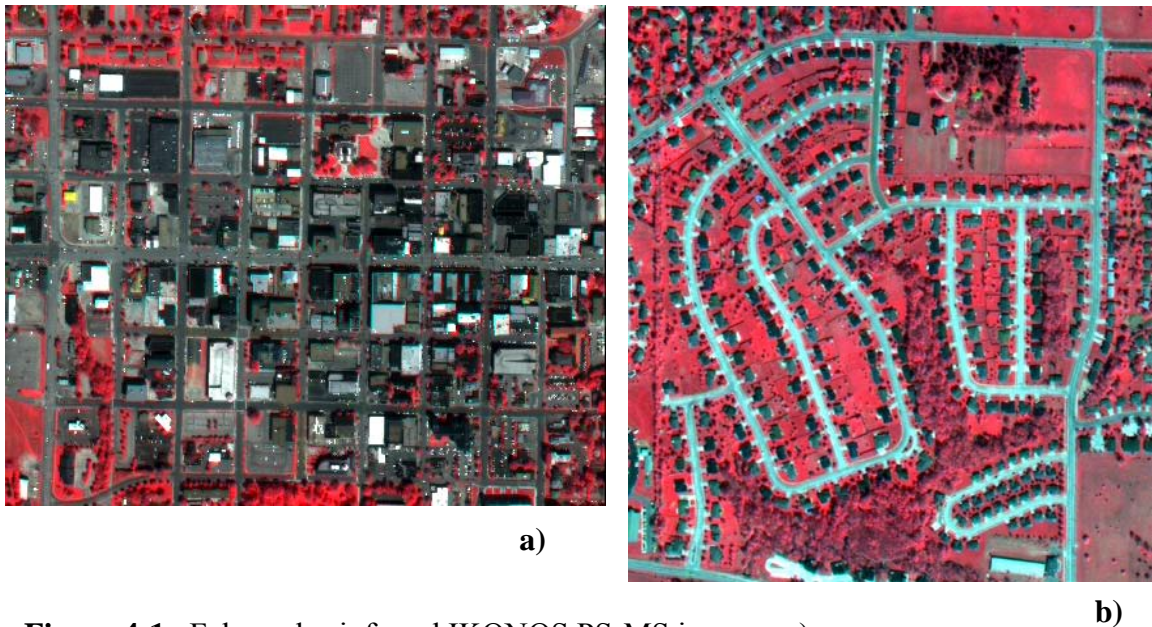
where  $L_R$  is the total length of the reference road network,  $L_E$  is the total length of the road extraction, and  $L_{UR}$  is the length of the unextracted reference road network. The completeness statistic is the percentage of the reference road network which has been extracted, and is a measure of producer's accuracy. The correctness statistic is the percentage of the extracted road network which falls within the reference buffer, and is a measure of user's accuracy. Quality is a composite measure providing an overall assessment of the accuracy of the extracted road network. By using three road extraction accuracy measures, we are able to get a complete picture of the performance of the road extraction algorithm. For example, a completeness value of 1.0 can be obtained by simply identifying the entire image as a road, however this will result in many false positives and a very low correctness value. Conversely, a correctness value of 1.0 can be

obtained by correctly extracting only a single road, however most of the road network will be unextracted, resulting in a very low completeness value. The quality measure is useful as an overall accuracy measure to compare the results of several different road extractions.

### **4.3 Road Extraction from Land Cover Classification**

Due to spectral variation in the roads and noise present in the imagery, automated road network extraction directly from digital imagery is a difficult task. The problem can be simplified by extracting the roads from a classification image. The road network extraction techniques examined in this section follow this approach. For the road network extraction to be successful, an accurate classification of the imagery is necessary. The classification used here for road extraction is the combined fuzzy pixel and object-based approach described in the previous chapter. Once an accurate urban land cover map has been generated, the road network and road centerlines can be extracted. Two different techniques for road extraction from a land cover classification image are presented in this section. The first technique examined was to simply skeletonize the *Road* class pixels. This approach was found to produce a large number of false positive responses, and does not have the ability to identify roads in areas where there are errors in the initial classification. The second road extraction technique examined was an iterative identification and growth process where road segments are identified and added to the road network, beginning with the longest roads in the imagery and subsequently adding shorter length road segments as the algorithm progresses. This technique greatly reduces the number of false positives and has the ability to identify road

segments in areas where there are errors in the initial classification. To analyze the performance of these two techniques, two test sites, containing both urban and suburban land cover, were chosen from the IKONOS image dataset of Columbia. The test sites are the same as those shown in Figs. 3-4 and 3-9. Subsets of the test sites are shown in Fig. 4-1.



**Figure 4-1.** False color infrared IKONOS PS-MS images: a) subset of urban test site, and b) subset of suburban test site.

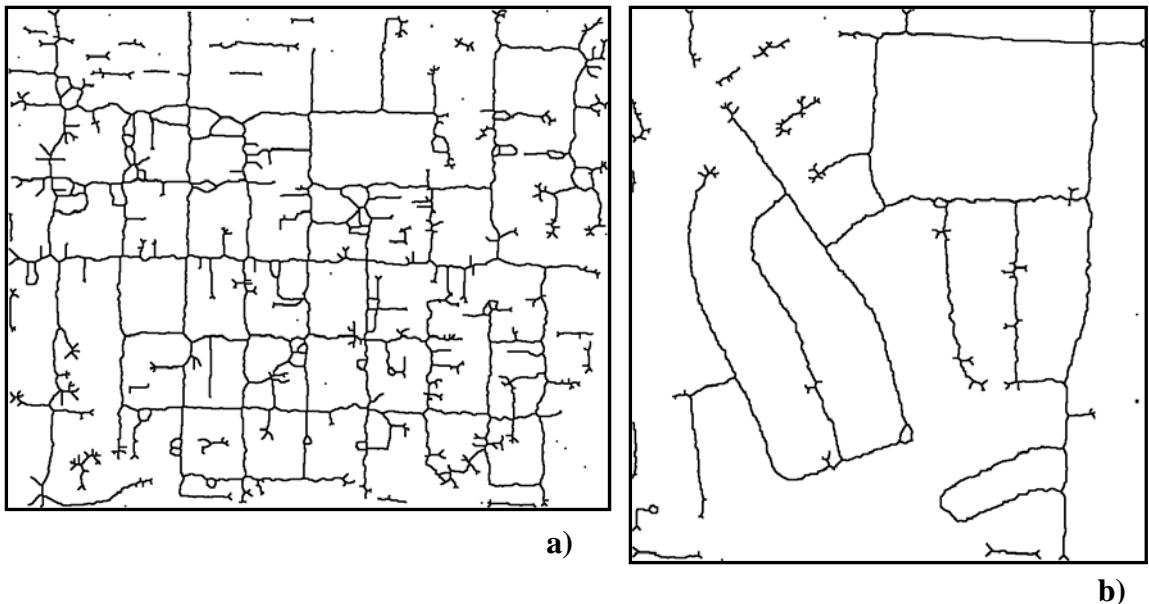
#### *A. Skeletonization*

The initial technique used for extraction of road centerlines from an urban land cover classification is based on morphological thinning of the *Road* pixels to produce a skeleton. The skeletonization algorithm described in Section 3.3-B is used. After producing the skeleton from morphological thinning, it was pruned to help decrease the number of spurs in the skeleton. Pruning is a morphological technique that removes any line segments shorter than a predefined length from the skeleton [Gonzalez and Woods,

2002]. Skeletonization and pruning was applied to the land cover maps of both the urban and a suburban test sites. The road extraction performance measures calculated from the skeletonization are shown in Table 4-1. Subsets of the road centerlines extracted from the test sites are shown in Fig. 4-2. Most of the road centerlines are correctly identified, however there are significant amounts of non-road objects identified as road centerlines. The completeness measures for the dense urban area and the suburban area are very similar, however the correctness measure for the dense urban area is significantly lower than for the suburban area. This is because there are significantly more non-road objects identified as road centerlines in the urban area test site than in the suburban test site.

**Table 4-1**  
**Road Extraction Performance from Skeletonization and Pruning**

	<i>Completeness</i>	<i>Correctness</i>	<i>Quality</i>
Urban	0.759	0.537	0.476
Suburban	0.764	0.755	0.641



**Figure 4-2.** Road network extracted from land cover classification by skeletonization. a) subset of urban test site (Fig. 4-1 a), and b) subset of suburban test site (Fig. 4-1 b).

### *B. Iterative Road Segment Identification and Growth*

As seen in Table 4-1 and Fig. 4-2, the skeleton of the *Road* pixels overestimates the road network resulting in false positives and a poor value for the correctness measure, especially in the urban test site. A more accurate road network can be obtained using an iterative process to extract and grow the road centerlines from the urban land-cover map. Roads are identified as long linear segments of *Road* pixels, with longer segments having a greater confidence of belonging to the road network than segments with short length. Long line segments are iteratively added to the road network and a buffer is set up around them to exclude any line segments that are similar in angle to the identified road network segments. This helps to avoid overestimation of the road network. As line segments are added to the road network, the algorithm examines the endpoints of the line segments and, using parameters set by the user, allows the newly added line segment to grow through small gaps and around corners in the road network. The algorithm continues to iterate, adding new line segments to the road network, until no line segments larger than the minimum length set by the user are available.

Before the extraction of road segments can begin, linear segments of *Road* pixels must first be identified. The classification image is converted to a binary image with the pixels classified as *Road* set to '1' and all other classes to '0'. Line segments are then identified by applying the 2-D spatial contextual feature extraction algorithm, described in Section 2.3-B, to the binary classification image. The road centerlines are then extracted, as summarized above. For a detailed description of the iterative identification and growth of road segments from the spatial features, see Section 4.4-B below.

The road extraction performance measures obtained when applying this algorithm

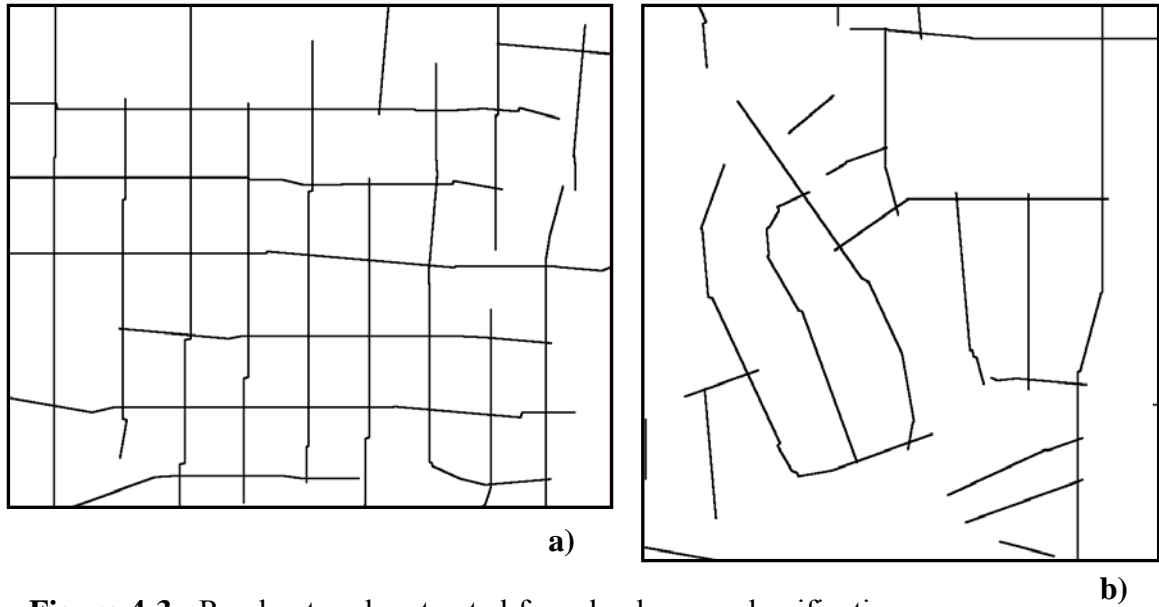
to the two test sites are shown in Table 4-2. A subset of the road network obtained in the dense urban area using the iterative algorithm is shown in Fig. 4-3a. As can be seen in Figure 4-3 and by examining the correctness measures, the iterative road network extraction algorithm removes most of the false positives and produces an accurate road network. Consequently the correctness values increase to 87-93%, which is a significant improvement over the results in Table 4-1 based on skeletonization of the *Road* pixels from the initial land cover map. However, the completeness measures for both the urban and suburban test sites decrease by about 5% compared to the results in Table 4-1. The completeness values of approximately 70% indicate that a significant amount of the roads are presently not being detected. There is an increase in the quality measure of 6–17%, indicating an improvement in overall accuracy for both test sites.

**Table 4-2**  
**Road Extraction Performance Measures from Iterative Algorithm**

	<i>Completeness</i>	<i>Correctness</i>	<i>Quality</i>
Urban	0.705	0.868	0.643
Suburban	0.708	0.933	0.702

#### **4.4 Road Extraction from High Resolution Imagery**

In the previous sections, several techniques for extracting road centerlines from land cover classifications were presented. The iterative road extraction technique was able to produce an accurate but incomplete road network. In this section, a fully automated road network extraction algorithm is presented. The fully automated method extracts the roads directly from the imagery and seeks to produce a more complete road



**Figure 4-3.** Road network extracted from land cover classification by iterative algorithm. a) subset of urban test site (Fig. 4-1 a), and b) subset of suburban test site (Fig. 4-1 b).

network than the results in the previous sections while maintaining a high level of accuracy.

While the appearance of roads in high-resolution imagery varies widely, they display several common characteristics that can be exploited for fully automated extraction. Regardless of their context, all roads have the same general shape, that of a long narrow ribbon that may or may not be curved. In contrast to imagery obtained from medium-resolution sensors (e.g. Landsat) where roads have a width of only a single pixel, the width of the roads is measurable in high-resolution imagery and can be used to aid extraction. Roads are also made up of non-vegetation pixels, which are easily identified when working with multispectral imagery. In addition, the topology of the road network can be used to aid in extraction and reduction of false positive road identification. Roads running parallel to each other in a road network are usually spaced a minimum distance apart, i.e. the width of a city block for example. If two identified roads are nearly parallel



and very close together, one of them is likely to be a false positive road identification. Another property of road networks is that the angle between two intersecting roads is typically close to  $90^\circ$  and very rarely is the intersecting angle less than  $45^\circ$ . By only allowing roads to be added to the network that have near perpendicular intersections, a large number of false positive road identifications can be avoided.

A road is initially identified as a connected group of spectrally similar non-vegetation pixels oriented in a long narrow rectangular shape. Therefore, two types of information must be extracted from the imagery before the road network identification can begin: (1) the 2-D spatial extent of spectrally similar groups of connected pixels, and (2) the vegetation status of those pixels. Length-width contextual features are used to quantify the 2-D spatial extent of objects in the imagery, and a vegetation index is used for identification of vegetation pixels. The initial identification only identifies straight road segments. Curvilinear roads are modeled as a group of linear segments connected together in piecewise linear fashion. Although initially only the linear portion of a curved road is identified, in subsequent processing the curved portion of the road can also be extracted (see Section 4.4-B).

#### *A. Feature Extraction*

While many structures in the urban environment are spectrally similar to roads, such as buildings and other impervious surface, roads are typically oriented along a long narrow ribbon while other urban land cover types tend to have a more compact shape. In Chapter 2, a spatial contextual measure was used to increase discrimination between urban land cover types with significant spectral confusion/overlap in a fuzzy

classification scheme. The spatial measure examines the context of each pixel and provides quantitative estimates for the 2-D spatial extent and directionality of groups of spectrally similar connected pixels. These quantitative measures are computed for each pixel and they form the basis for the road network extraction algorithm presented in Section 4.4-B. The extraction of the 2-D spatial contextual features is described in detail in Section 2.3-B.

An advantage of working with multispectral imagery, as opposed to data with only a panchromatic channel, is the ease with which vegetation can be identified. Vegetation in the image is identified and masked out using the normalized difference vegetation index (NDVI) [Jenson, 1996]. The NDVI is calculated for each pixel using

$$NDVI = \frac{NIR - R}{NIR + R} \quad (4-4)$$

where  $NIR$  and  $R$  are the near-infrared and red band channel data, respectively. After the NDVI is calculated, vegetation pixels are identified as those pixels that exceed a threshold of 0.2.

#### *B. Road Network Extraction*

The road network extraction algorithm presented here is an iterative process that first identifies and then grows road segments using the features discussed in the previous section and knowledge of the general characteristics and topology of a road network. Roads are initially identified as long linear segments of non-vegetation pixels, with longer segments having higher confidence of being part of the road network than segments with short length. The algorithm begins by examining the longest length line segment, e.g.  $L_j^{max}$ , present in the imagery, progressing to smaller length line segments as

it iterates. Once a potential road segment has been identified, the algorithm examines the endpoints of the line segment, and using parameters set by the user, attempts to track the segment through small gaps and around curves in the road network. As road segments are iteratively added to the road network, a buffer is set up around them to exclude any line segments that are similar in angle and close to previously identified road network segments. This helps avoid overestimation of the road network and also eliminates multiple responses originating from a single road. The algorithm continues to iterate, adding new line segments to the road network until no line segments larger than the minimum length set by the user are available. These steps are described in more detail in the following subsections. Figure 4-4 presents a processing flow chart summarizing the steps of the road extraction algorithm.

### *1. Identification of Road Line Segments*

At the beginning of each iteration, the longest line segment still available in the image is identified and then examined to determine if it matches the criteria for a road. (In the first iteration, this is the longest line segment present in the imagery.) The pixel  $j$  (non-vegetation) with the maximum line segment length value,  $L^{max}_j$ , is located and the corresponding line segment is identified using  $\theta_j$  and the endpoints,  $EP^1_j$  and  $EP^2_j$ . The criteria a line segment must meet to be identified as a potential road segment are: 1) the average of the minimum length values,  $L^{min}$ , for all pixels in the line segment must be less than the road width threshold,  $L_{wdh}$ ; 2) the majority of the angle values,  $\theta$ , of the pixels along the line segment must be in agreement with  $\theta_j$ ; and 3) the line segment must be comprised of non-vegetation pixels. If the line segment does not meet these road criteria,

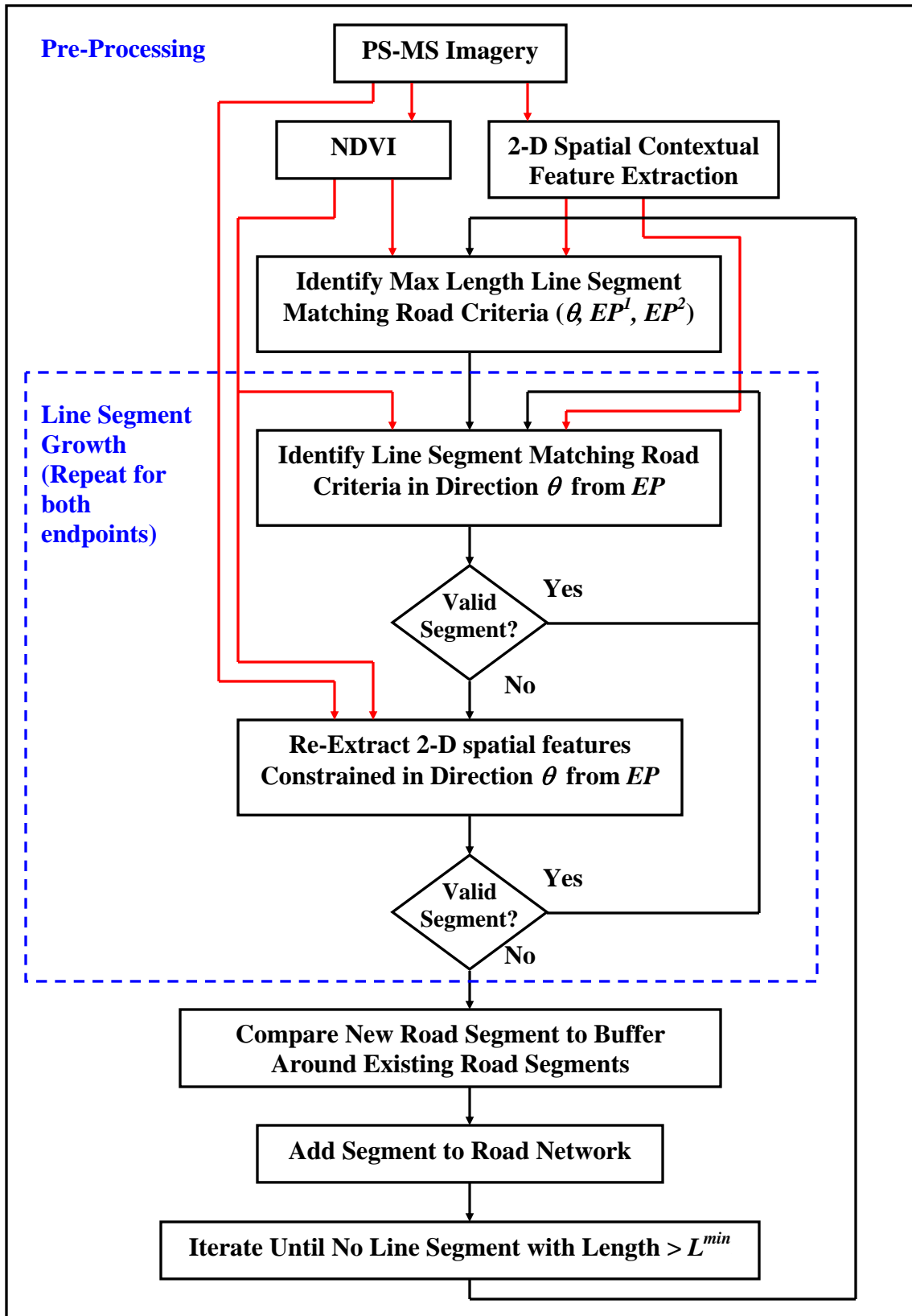


Figure 4-4. Processing flow chart summarizing the steps of the road extraction algorithm.

then  $L^{max}$  for the pixels in the line segment are set to zero so these pixels will not be identified again as a potential road segment. This step is repeated until a line segment matching the road criteria is found.

## 2. Line Segment Growth

Once a line segment satisfying all road criteria has been identified, the algorithm attempts to grow the ends of the line segment so roads can be tracked around curves and through areas partially occluded by trees and/or shadows. Two techniques are employed to grow the line segment ends. First, the 2-D spatial features in the areas around the endpoints of the line segment,  $EP^1_j$  and  $EP^2_j$ , are examined and new line segments in those areas that meet the road criteria and extend in the same general direction as  $\theta_j$  are identified. If no such line segment can be found, the length feature is re-extracted from the points around the endpoints of the line segment. Unlike the feature extraction phase where the search directions are evenly spaced in all directions, here the length extraction is constrained to only search along lines in the same general direction of  $\theta_j$ . The longest line segment identified by the constrained re-extraction of length values is then subjected to the road criteria to make sure a valid segment has been identified. The two growth techniques are repeated for both endpoints until no line segments can be identified and added to the road segment. The growth phase of the algorithm is controlled by the parameter  $\theta_{grw}$ , the maximum angle threshold between the line segment and new segments added during the growth phase.  $\theta_{grw}$  should be set large enough such that curved roads are tracked but not so large that the extracted roads are tracked around corners and onto different roads. The extracted road segment with extended endpoints is

identified as  $RS_j$ . The line segment length values,  $L_{max}$ , for all pixels in  $RS_j$  are set to zero so no portion of  $RS_j$  will be analyzed in subsequent iterations of the algorithm.

### 3. Road Buffer Examination

As road segments are iteratively identified and added to the road network, a buffer is defined around each road segment. This buffer enforces the topology of road networks discussed above, restricting roads to intersecting each other at near perpendicular angles and specifying a minimum separation between near parallel roads. An additional effect of the buffer is to give preference to longer length line segments being added to the road network. Road segments with large length values are identified and buffered early in the iteration of the algorithm, preventing shorter line segments, which have a greater chance of being in error, from being added to the network near longer length segments in ways that violate the structure of the road network. The effects of the buffer are controlled by the width of the buffer,  $L_{buff}$ , and the minimum angle between road segments threshold,  $\theta_{buff}$ . Before the road segment  $RS_j$  can be added to the extracted road network, it must be compared with the buffer surrounding the previously extracted road network to determine if  $RS_j$  is in conflict with any of the previously extracted roads. If  $RS_j$  falls inside the buffer of a previously extracted road segment, the angle of  $RS_j$ , i.e.  $\theta_j$ , is compared to the angle of the buffered road segment. If the angle between  $RS_j$  and the previously extracted road network segment is less than  $\theta_{buff}$ , then the portion of  $RS_j$  falling within the buffer will not be added to the road network. The sections of  $RS_j$  that are not in conflict with the buffer are added to the extracted road network, and the buffer is modified to reflect the newly added road segment. The algorithm now cycles to a new iteration, going through

the steps described above until no line segment length values,  $L^{max}$ , are found larger than the minimum line segment length,  $L_{stop}$ .

If the buffer size,  $L_{buff}$ , and the minimum allowable length,  $L_{stop}$ , are set properly, false positive road identification in dense urban areas can be avoided. Generally speaking,  $L_{buff}$  should be chosen to be slightly less than the typical size of city blocks in the road network and  $L_{stop}$  should be larger than most city blocks in the road network. This allows only line segments at least as long as a city block to be examined for addition to the road network and keeps linear features in large commercial structures from being identified as part of the road network. An example illustrating the extraction of a curved road segment is shown in Figure 4-5.

### *C. Test Results*

#### *1. IKONOS Imagery*

The road extraction algorithm described in this section was first tested on two sites within the IKONOS Columbia imagery, one with urban land cover and one with suburban land cover. The urban test site (Fig. 3-4) measures 2000 x 3000 m and the suburban test site (Fig. 3-9) measures 1000 x 2428 m. Subsets of the extracted road network from these two test sites are shown in Fig. 4-6. The road extraction performance measures obtained after applying the algorithm to the two test sites are shown in Table 4-3. The algorithm parameters are shown in Table 4-4. The minimum line segment length,  $L_{stop}$ , was the only parameter changed between the urban and suburban test sites. A smaller value was used for the suburban test site to better track curvilinear roads. Almost all roads are at least partially extracted, however there are some gaps in the road network.



**Figure 4-5.** Illustration of road segment growth: a) identification of initial road segment, b) examination of length-width values around endpoint 1, c) resulting line segment, d) examination of length-width values around endpoint 2 (no valid line segment identified).





e)



f)



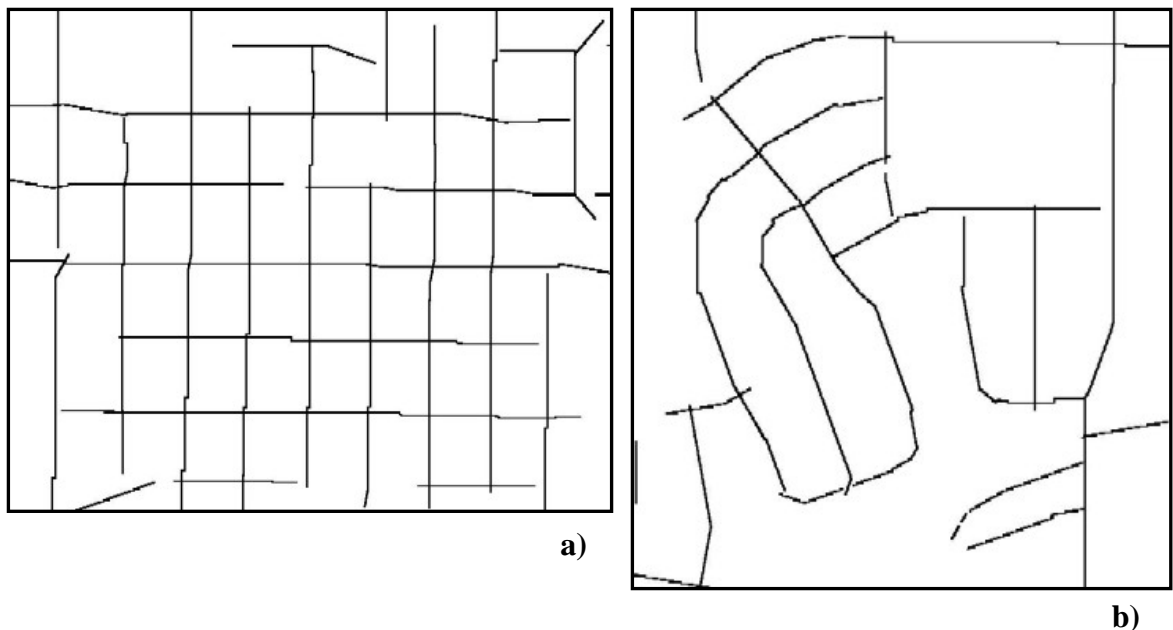
g)



h)

**Figure 4-5 cont.** Illustration of road segment growth: e) re-extraction of length value, constrained to direction of line segment, f) resulting line segment, g) final result of road segment growth, and h) new road segment compared to buffer of previously extracted road segments.

The suburban test site has a completeness measure that is 8% smaller compared to the urban test site, and this is due to the presence of a large number of curvilinear roads in the suburban test site. However, the correctness measure for the urban test site is 17% smaller than that of the suburban test site. This decrease in accuracy in the dense urban area is expected since there are many more opportunities for error in these types of areas due to the large numbers of non-road linear features. There is an increase of 17% in completeness and a decrease of 16% in correctness for the urban test site when the fully automated technique is compared to the iterative technique for extracting roads from a classification. The suburban test site shows an increase of 8% in completeness and a decrease of 6% in correctness when the two techniques are compared. In both test sites, the increase in completeness is slightly greater than the decrease in correctness. When comparing the quality measure for the two techniques, there is a small increase for the



**Figure 4-6.** Road network extracted from IKONOS imagery by fully automated algorithm. a) subset of urban test site (Fig. 4-1 a), and b) subset of suburban test site (Fig. 4-1 b).

suburban test site, while this measure is essentially the same for the urban test site. It should be noted that while both techniques yield approximately the same value of quality, the first technique operates on a classification image, requiring training data, while the second road extraction technique is fully automated.

**Table 4-3**  
**Road Extraction Performance Measures for IKONOS Test Sites**

	<i>Completeness</i>	<i>Correctness</i>	<i>Quality</i>
Urban	0.872	0.704	0.642
Suburban	0.791	0.873	0.730

**Table 4-4**  
**Road Network Extraction Algorithm Parameters used for IKONOS Test Sites**

	$L_{wdh}$	$\theta_{grw}$	$L_{buff}$	$\theta_{buff}$	$L_{stop}$
Urban	20 m	30°	121 m	60°	100 m
Suburban	20 m	30°	121 m	60°	80 m

## 2. *Digital Aerial Orthophotos*

The road network extraction algorithm was also tested on multispectral digital aerial orthophotos of St. Charles, MO acquired in February 2002. The digital orthophotos are part of a joint program between NIMA/NGA and the USGS to acquire high-resolution digital imagery for 133 urban cities in the U.S (see Section 1.2-A). The digital orthophotos of St. Charles were originally acquired at 8.3-inch resolution, but these were resampled to 1 m so the results could be compared to those from the IKONOS imagery. Because the digital orthophotos do not have a near infrared band, the NDVI cannot be used for vegetation identification. To identify the vegetation in this imagery,

the data were classified using an unsupervised clustering technique [Theodoridis and Koutroumbas, 1998], and the vegetation classes were manually identified. While this step does require more interaction from the user, it is a fast and simple process and can be done using many commercial image processing software packages.

Two test sites within the St. Charles dataset were selected for road network extraction, a site with commercial/industrial landcover and a site with suburban landcover. The commercial/industrial test site measures 1200 x 1800 m and the suburban test site measures 1000 x 1000 m. These two test sites and the extracted roads are shown in Figs. 4-7 and 4-8. The road extraction quality measures obtained after applying the algorithm to the two test sites are shown in Table 4-5. The algorithm parameters are shown in Table 4-6. The parameters controlling the buffer,  $L_{buff}$  and  $\theta_{buff}$ , for the commercial/industrial test sight have been greatly relaxed to better extract the multi-lane highway passing through the center of the test area. When the accuracy measures for the commercial/industrial test site are compared to those obtained from the suburban test site, they show the same trends as those obtained from the IKONOS imagery, although the variation is somewhat reduced. The road extraction accuracy measures are reduced from those obtained from the IKONOS imagery, possibly due to the lack of a near infrared band in the data set.

**Table 4-5**  
**Road Extraction Performance Measures for Digital Aerial Orthophotos**

	<i>Completeness</i>	<i>Correctness</i>	<i>Quality</i>
Commercial / Industrial	0.804	0.705	0.617
Suburban	0.791	0.843	0.712



**Table 4-6**  
**Road Network Extraction Algorithm Parameters used for Digital Aerial**  
**Orthophotos**

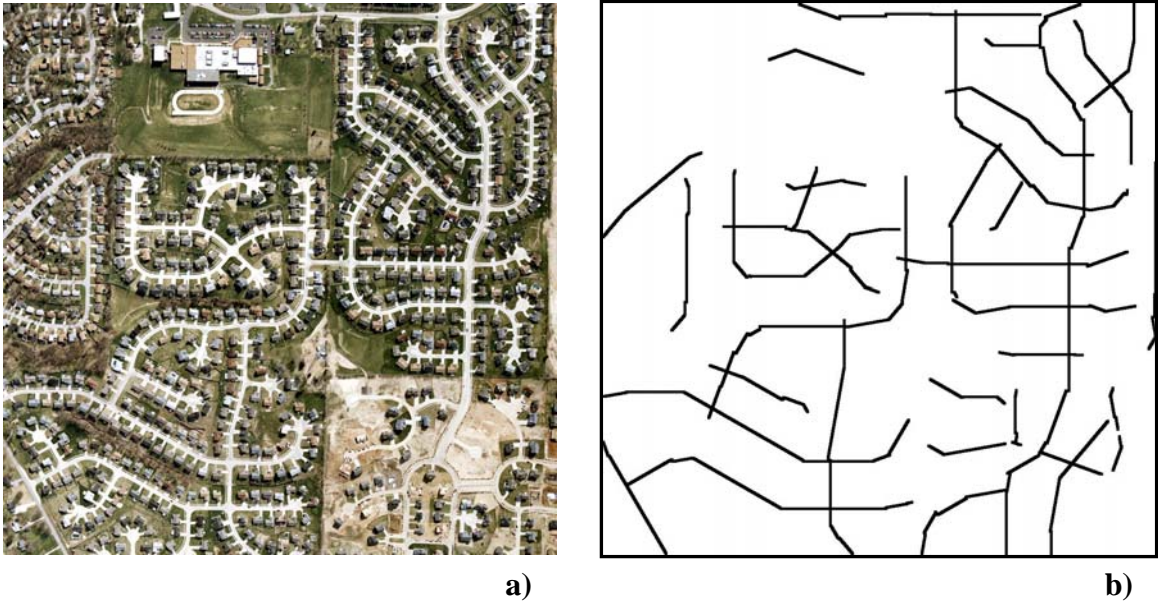
	$L_{wdh}$	$\theta_{grw}$	$L_{buff}$	$\theta_{buff}$	$L_{stop}$
Industrial	20 m	30°	31 m	45°	100 m
Suburban	20 m	30°	101 m	60°	100 m



**Figure 4-7.** Road network extracted from digital orthophoto by fully automated algorithm. a) commercial/industrial test site, b) extracted road network.

#### 4.5 Summary

The problem of automated extraction of the road network from high-resolution imagery over urban areas was examined in this chapter. The techniques presented here seek to extract a single pixel wide response for each road, providing an estimate of the



**Figure 4-8.** Road network extracted from digital orthophoto by fully automated algorithm. a) suburban test site, b) extracted road network.

road centerline. Two techniques for extracting roads from land cover classification maps were presented: skeletonization and an iterative line segment identification and growth algorithm. Skeletonization was found to produce large numbers of false positive road identifications. The road network extracted by the iterative algorithm was accurate (87-93% correctness), however it was only able to produce a partial extraction resulting in completeness values of 70%.

Following the techniques for road extraction from land cover classifications, a fully automated road network extraction algorithm for high-resolution multispectral data over urban areas was presented. Roads were extracted using 2-D spatial feature information after masking out vegetation pixels using the NDVI statistic. An iterative road segment identification and growth algorithm was then introduced. This algorithm iteratively adds road segments to the road network, beginning with the longest length line

segments in the imagery and progressing to segments of shorter length. The endpoints of the road segments are grown so curvilinear roads can be extracted. A buffer around the road network is utilized to keep spurious line segments from being added to the existing road network. The extraction algorithm was tested on several sites from an IKONOS high-resolution satellite image and digital aerial orthophotographs. The results from both datasets show that the algorithm is more accurate in suburban areas, where there is a reduced chance for error. However, the algorithm produces a more complete road network in urban areas where the presence of curved roads is minimal.

The material presented in this chapter was presented at the *2003 GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas* [Shackelford and Davis, 2003a], and the *2003 International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2003b]. The main contribution of this work to the body of literature on road extraction from high-resolution multispectral imagery of urban areas is the development of a fully automated road network extraction technique with the ability to produce an accurate and complete road network. The algorithm was shown to produce road networks that were 79–87% complete and 70–87% correct. This compares favorably with results reported for other techniques where road network completeness values of 49%–87% with correctness values of 80%–95% have been reported [Haverkamp, 2002], [Hinz and Baumgartner, 2003], [Shi and Zhu, 2001], [Dell’Acqua and Gamba, 2001]. In particular, the 79%–87% road completeness measures from this work are significantly better than the 49%–67% completeness results reported thus far for other road extraction techniques applied to high-resolution commercial satellite imagery

[Haverkamp, 2002], [Shi and Zhu, 2001]. A comparison between the results presented here and those reported in the literature is shown in Table 4-7.

Author(s)	Year	Data Source	Pixel Size	Automated	Completeness (%)	Correctness (%)	Quality (%)	Comments
Shackelford & Davis	2003	Satellite	1 m	Semi	71	97-93	64-70	Urban + Suburban
Shackelford & Davis	2003	Satellite	1 m	Full	79-87	70-87	62-73	Urban + Suburban
Shi & Zhu	2002	Satellite	1 m	Full	92	-	-	Urban - Straight line only
Soille & Pesaresi	2002	Satellite	5 m	Full	-	-	-	Urban
Haverkamp	2002	Satellite	1 m	Full	49-66	-	-	Urban + Suburban
Couloigner & Ranchin	2000	Satellite	2 m	Semi	-	-	-	Urban
Doucette <i>et al.</i>	2001	Satellite / Aerial	1 m	Semi	-	-	-	Urban + Suburban
Agouris <i>et al.</i>	2001	Satellite / Aerial	1 m	Full	-	-	-	Urban + Suburban
Hinz & Baumgartner	2003	Aerial DSM	15 cm 1 m	Semi	75-85	95	-	Urban
Baumgartner <i>et al.</i>	1999	Aerial	0.2- 0.5 m	Semi	80-90	90-95	-	Rural
Stegar	1998	Aerial	1 m	Full	-	-	-	Rural
Mayer <i>et al.</i>	1997	Aerial	0.2 – 0.5 m	Semi	-	-	-	Rural
Dell'Acqua & Gamba	2001	SAR	7.5 m	Full	72-82	62-86	54-64	Urban

**Table 4-7.** Comparison of results reported in automated road extraction literature.



## Chapter 5

### Automated 2-D Building Footprint Extraction

#### 5.1 Introduction

The two most prominent features characterizing the urban environment are the presence of a road network and urban buildings. In the previous chapter, a fully automated algorithm for road network extraction from high-resolution multispectral (MS) imagery was presented. Here, we present a fully automated algorithm for two-dimensional building footprint extraction. As discussed in Chapter 1, there are many applications for these features, such as urban growth assessment, infrastructure planning, emergency management and homeland security applications. Although there are a number of algorithms for automated extraction of three-dimensional building structures from high-resolution digital elevation data, the extraction of 2-D building footprints from high-resolution imagery is done mostly by manual digitization, and this is both time consuming and expensive.

Buildings in urban areas normally have significant variations in shape, size, and spectral response, thereby making it difficult for any single technique to detect all building types in an image. For this reason, a multi-detector fusion strategy is adopted for automated building extraction. Two distinct building detectors are developed, each with complementary strengths and weaknesses, to produce estimates of 2-D building shape, size, and position. While spatial and spectral building appearance in high-

resolution imagery can vary significantly, there are several characteristics common to most buildings, such as rectilinear shape and cast shadows, which can be exploited for automated detection. Both rectilinear shape and proximity to shadows were used by the object-based fuzzy logic classifier for discrimination between buildings and impervious surface (see Chapter 3). However that approach was supervised and therefore required human-generated input training data. Here, our goal is to develop a fully automated building extraction technique. Thus, only processing techniques that require no training data are applicable. Several of the techniques used in the object-based classifier, such as segmentation and fuzzy shape characterization, are unsupervised and are used here for automated building extraction. In addition to segmentation and fuzzy shape analysis, buildings and shadows are also identified through the use of a multi-scale morphological image analysis technique. Two building detectors are utilized: direct detection through shape analysis and indirect detection through identification of cast shadows. The shadow identification is followed by a shape analysis of objects generated from the segmented pan-sharpened MS (PS-MS) imagery.

## **5.2 Differential Morphological Profile**

The building detectors developed here attempt to identify buildings and their shadows from a differential morphological profile (DMP) of the high-resolution panchromatic (PAN) image data. Mathematical morphology employs a set of image operators to extract and analyze image components based on the shape and size of quasi-homogeneous regions in the image. Mathematical morphology has been applied to a wide variety of practical problems such as image pre-processing, noise filtering, shape

detection and decomposition, and pattern association. The DMP is a relatively new technique, introduced by Pesaresi and Benediktsson [2001] where it was used to segment high-resolution satellite imagery. The DMP was also used for urban land cover classification by Benediktsson *et al.* [2003]. The DMP is a multi-scale image analysis technique where a morphological profile of the image is constructed through the use of morphological opening and closing by reconstruction operations while varying the size of the structuring element (SE).

#### *A. Morphological Opening and Closing by Reconstruction*

The DMP is constructed using morphological opening and closing by reconstruction operators, which differ significantly from the traditional morphological opening and closing operators. Traditionally, the opening  $\gamma$  of an image  $I$  with  $SE=\lambda$  is defined as an erosion  $\varepsilon_\lambda$  followed by a dilation  $\delta_\lambda$  using the same SE. The closing  $\varphi$  of  $I$  with  $SE=\lambda$  is defined as a dilation  $\delta_\lambda$  followed by an erosion  $\varepsilon_\lambda$ .

The opening by reconstruction  $\gamma^*$  of an image  $I$  with  $SE=\lambda$  is defined as the erosion of  $I$  with  $SE=\lambda$  followed by grayscale geodesic dilation iterated until stability is reached.  $\gamma^*_\lambda(I)$  can be calculated using the following steps:

$$\begin{aligned}
 &J(0) = \varepsilon_\lambda(I) \\
 &n = 0 \\
 &\text{Repeat} \\
 &\quad J(n+1) = \delta_1(J(n)) \wedge I \\
 &\quad n = n+1 \\
 &\text{Until } J(n+1) = J(n) \\
 &\gamma^*_\lambda(I) = J(n+1)
 \end{aligned}$$

where  $\delta_1$  is dilation with SE of size one, and  $\wedge$  is the pointwise minimum. The effect of morphological opening by reconstruction is to remove all structures in the image that are

both smaller than the SE and brighter than their surroundings. Similarly, the closing by reconstruction  $\varphi^*$  of an image  $I$  with  $SE=\lambda$  is defined as the dilation of  $I$  with  $SE=\lambda$  followed by grayscale geodesic erosion iterated until stability is reached.  $\varphi^*_{\lambda}(I)$  can be calculated using the following steps:

$$\begin{aligned}
& J(0) = \delta_{\lambda}(I) \\
& \quad n = 0 \\
& \quad \text{Repeat} \\
& \quad \quad J(n+1) = \varepsilon_1(J(n)) \vee I \\
& \quad \quad n = n+1 \\
& \quad \text{Until } J(n+1) = J(n) \\
& \varphi^*_{\lambda}(I) = J(n+1)
\end{aligned}$$

where  $\varepsilon_1$  is erosion with SE of size one, and  $\vee$  is the pointwise maximum. The morphological closing by reconstruction operator removes all structures in the image that are both smaller than the SE and darker than their surroundings. For a more detailed description see Pesaresi and Benediktsson [2001] and Vincent [1993].

### B. Differential Morphological Profile

Here we present a review of the basic concept. Let  $\gamma^*_{\lambda}$  be a morphological opening by reconstruction operator with  $SE=\lambda$ . The opening profile at pixel  $x$  of image  $I$  is defined as a vector

$$\Pi\gamma(x) = \{\Pi\gamma_{\lambda} : \Pi\gamma_{\lambda} = \gamma^*_{\lambda}(x), \forall \lambda \in [0, n]\} \quad (5-1)$$

Let  $\varphi^*_{\lambda}$  be a morphological closing by reconstruction with operator  $SE = \lambda$ . Then, the closing profile at pixel  $x$  of the image  $I$  is defined as a vector

$$\Pi\varphi(x) = \{\Pi\varphi_{\lambda} : \Pi\varphi_{\lambda} = \varphi^*_{\lambda}(x), \forall \lambda \in [0, n]\} \quad (5-2)$$

In (5-1) and (5-2),  $\Pi\gamma_0(x) = \Pi\varphi_0(x) = I(x)$  for  $\lambda = 0$  by the definition of opening and closing by reconstruction. The derivative of the morphological profile is defined as a

vector storing the slope of the opening/closing profile for each step of increasing SE size.

The derivative of the opening profile is defined as the vector

$$\Delta\gamma(x) = \{\Delta\gamma_\lambda : \Delta\gamma_\lambda = |\Pi\gamma_\lambda - \Pi\gamma_{\lambda-1}|, \forall \lambda \in [1, n]\} \quad (5-3)$$

By duality, the derivative of the closing profile is defined as the vector

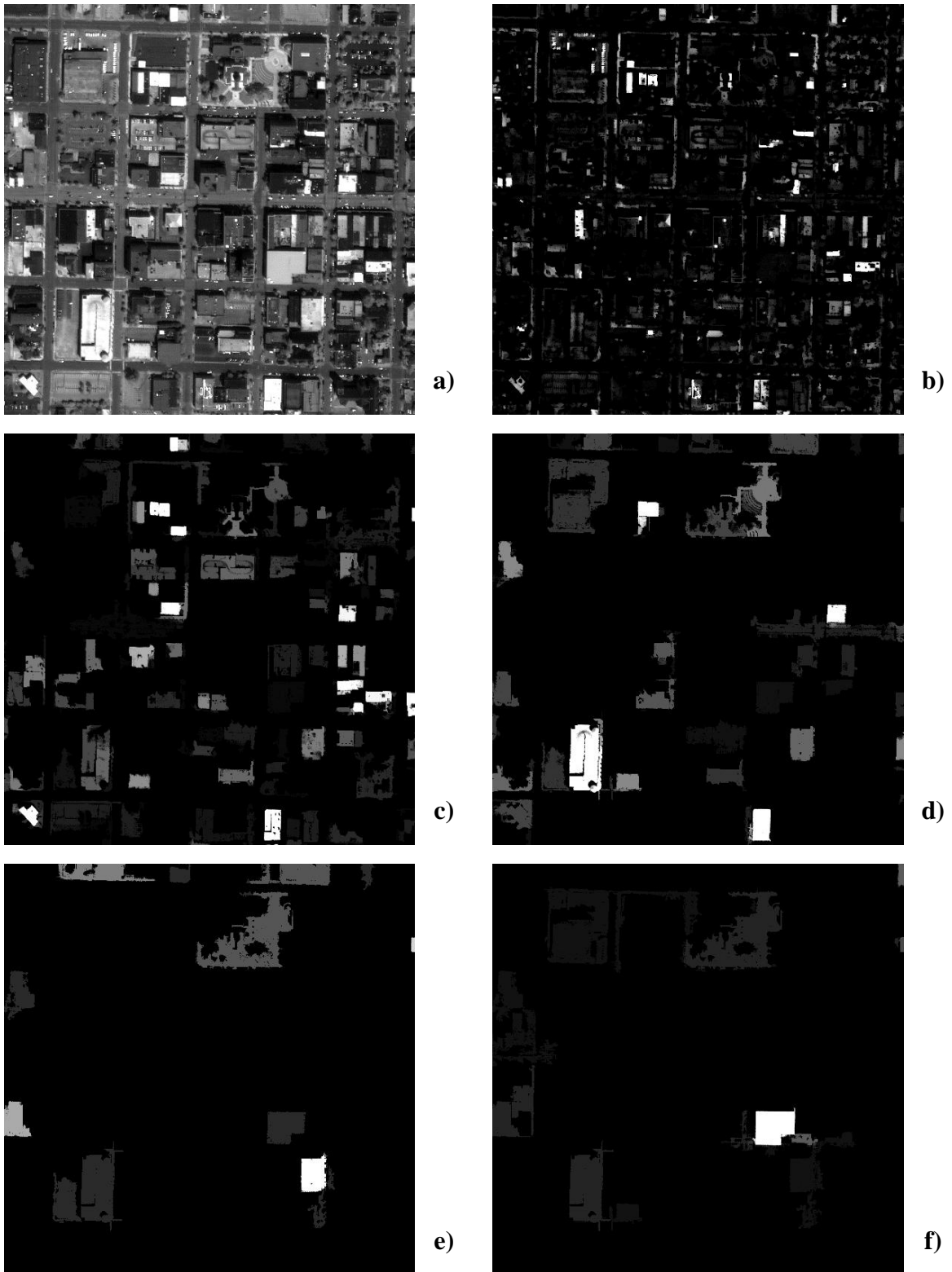
$$\Delta\varphi(x) = \{\Delta\varphi_\lambda : \Delta\varphi_\lambda = |\Pi\varphi_\lambda - \Pi\varphi_{\lambda-1}|, \forall \lambda \in [1, n]\} \quad (5-4)$$

In general, the derivative of the morphological profile (DMP) can be written as the vector

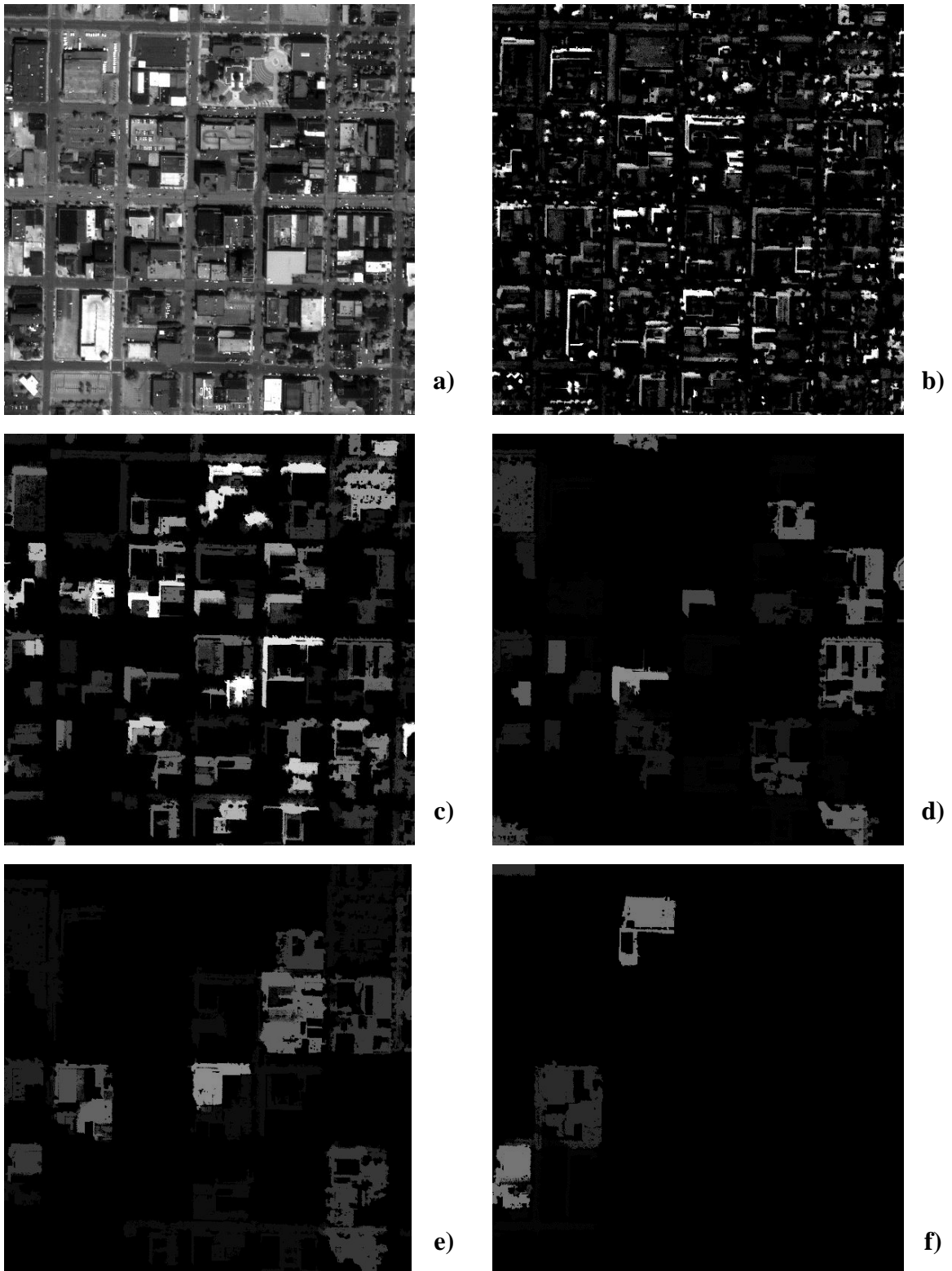
$$\Delta(x) = \left\{ \Delta_c : \begin{cases} \Delta_c = \Delta\varphi_{\lambda=n-c+1}, \forall c \in [1, n] \\ \Delta_c = \Delta\gamma_{\lambda=c-n}, \forall c \in [n+1, 2n] \end{cases} \right\} \quad (5-5)$$

where  $n$  is the total number of iterations (SE sizes), and  $c = 1, \dots, 2n$ . The responses of the derivative calculated using small SEs are near the central position of the DMP vector, while the responses for the largest SEs in the closing and opening profile are recorded at the beginning ( $c=1$ ) and the end ( $c=2n$ ), respectively.

The signal recorded in the DMP provides information about the size and contrast of multi-scale structures in the image. Small structures will have strong response for small SE sizes, yielding a large response near the center of the DMP, while large structures will have strong response for large SE sizes, yielding a large response near the beginning or end of the DMP. Structures darker than the surrounding background will have high response in the closing portion of the DMP ( $c = 1, \dots, n$ ), while brighter structures will have high response in the opening portion of the DMP ( $c = n+1, \dots, 2n$ ). An example of the DMP for a 1-m PAN image of an urban area is shown in Figs. 5-1 and 5-2 for the opening and closing operators, respectively. Disk shaped SEs are used, with radii,  $r$ , varying from 5 to 21 pixels, with a step size of 4 pixels. Note that many of the buildings present in the image are easily discriminated in the opening differential profile,



**Figure 5-1.** a) 1-m PAN image of dense urban area. Opening differential profile for radius b)  $r = 5$  m, c)  $r = 9$  m, d)  $r = 13$  m, e)  $r = 17$  m, and f)  $r = 21$  m.



**Figure 5-2.** a) 1-m PAN image of dense urban area. Closing differential profile for radius b)  $r = 5$  m, c)  $r = 9$  m, d)  $r = 13$  m, e)  $r = 17$  m, and f)  $r = 21$  m.

especially for SEs with  $r = 9$  and  $r = 13$ , while building shadows, which often appear as an “L-shape”, are easily seen in the closing differential profile.

### 5.3 Building Extraction

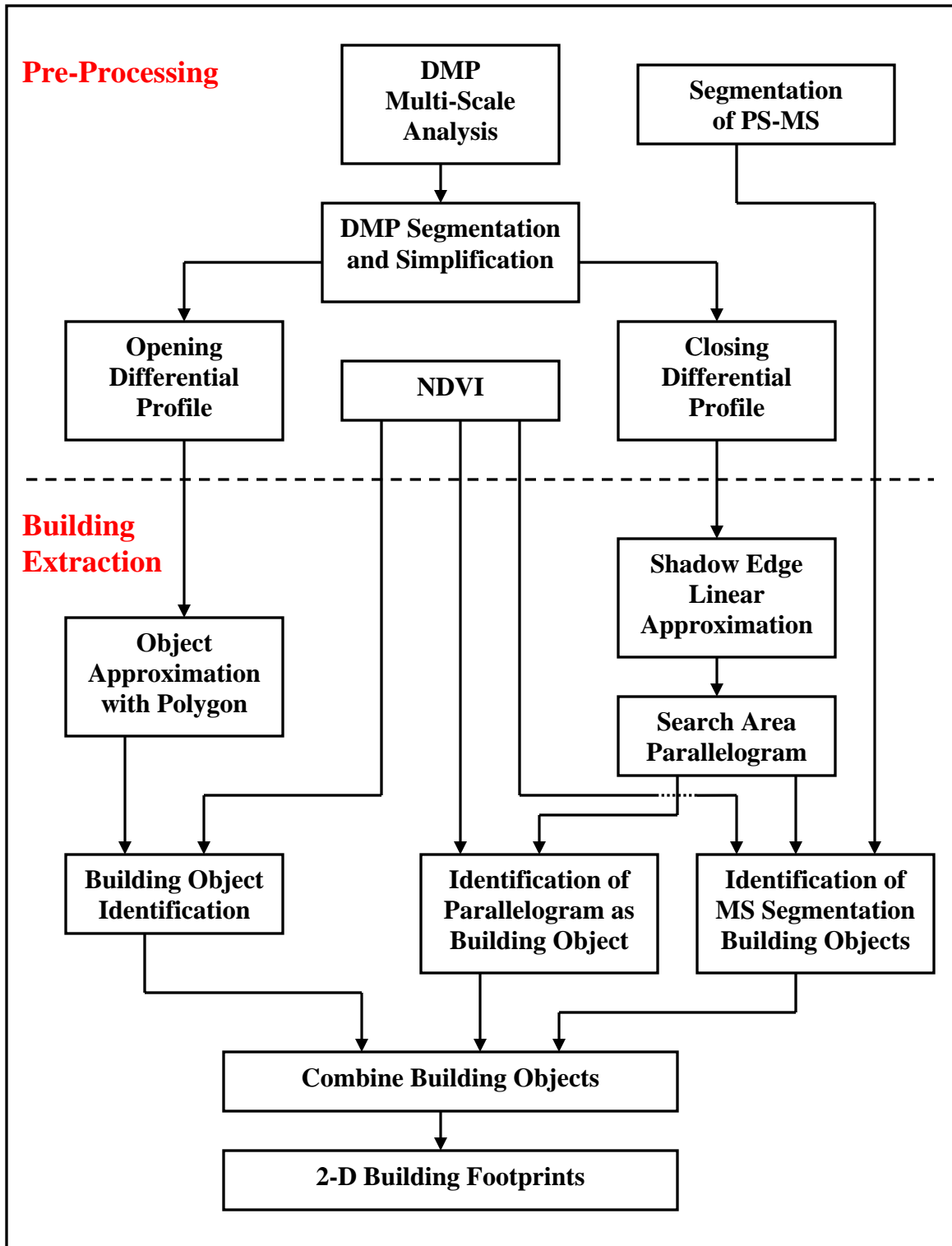
Due to the varying size of urban buildings, a multi-scale approach such as the DMP must be utilized for their extraction. In the DMP, structures with a similar scale (width) to that of the SE diameter will have a high response for that SE in the DMP, with bright structures in the opening portion of the profile and dark structures in the closing portion of the profile. As was shown in Figs. 5-1 and 5-2, many of the buildings and their shadows are readily visible in the DMP. For each pixel in the image, the position of the maximum response within the DMP vector,  $\Delta(x)$ , indicates both the SE size that best characterizes the structure that the pixel resides within and whether the pixel is part of a structure that is brighter or darker than the surrounding region. The actual value of the maximum DMP response indicates the magnitude of the contrast between the structure and its surrounding area. Regions with similar SE size, response, and contrast are delineated by segmenting each level of the DMP. Segments with dimensions significantly larger or smaller than the SE size of their level are rejected. After segmentation and removal of invalid objects, the redundancy between adjacent levels in the DMP is removed by merging overlapping objects, keeping objects from the opening portion of the DMP separate from those identified in the closing levels. This results in two sets of image objects: one for the opening portion of the differential profile and one for the closing portion. While all levels of the opening differential profile ( $r = 5$  m through  $r = 21$  m, step size of 4 m) are utilized for building extraction, only the  $r = 5$  m



and  $r = 9$  m levels of the closing differential profile are utilized as they are the only levels that provide useful shadow information. After delineation of image objects in the DMP it is possible to hypothesize candidate building and shadow objects. Buildings are identified from objects extracted from the opening differential profile, whereas building shadows are extracted from objects identified in the closing differential profile. The building shadows are used to hypothesize the presence of a building, which is then identified through shape analysis of objects extracted from a segmentation of the PS-MS imagery. Figure 5-3 presents a processing flow chart summarizing the steps of multi-detector building extraction algorithm.

#### *A. Building Extraction from the Opening Profile*

Several shape-based features, object dimensions and compactness, are utilized to identify buildings from among the objects extracted in the opening differential profile. Here, compactness is calculated as the ratio of the area of the object to the area of an approximating polygon. The shape of the approximating polygon is constrained to having six perpendicular sides. This type of polygon was chosen so that it could provide a reasonable approximation for a wide variety of building shapes, including both rectangular and “L-shaped” buildings. The first step in approximating the object with a polygon is to find the minimum area rectangle enclosing the object using the rotating calipers method [Toussaint, 1983]. A better approximation of the shape of the object is then found by identifying the corner of the minimum area rectangle furthest from the object and moving it as close as possible to the object while maintaining right angles between the edges of the approximating polygon, resulting in a six-sided polygon with



**Figure 5-3.** Processing flow chart summarizing the steps of the multi-detector building extraction algorithm.

right-angled corners. The ratio of the area of the object to the area of approximating polygon, a measure of rectilinear compactness, can then be calculated and used as a feature to identify building objects. Non-building objects with an amorphous shape have low compactness values, whereas buildings with quasi-rectangular shape have high compactness values. To avoid identifying objects that are either too large or too small to be buildings, the minimum length side of the enclosing rectangle and the area of the approximating polygon are also used as features. Finally, the percentage vegetation pixels present in the object, as determined from the NDVI, is utilized to keep from identifying rectilinear objects with primarily vegetative land cover (e.g. fields) as buildings.

#### *B. Shadow Extraction from the Closing Profile*

In addition to detecting buildings directly in the opening portion of the DMP, buildings are also indirectly identified by extracting their shadows from the closing differential profile. The extracted building shadows are used to define a search area where the presence of a building is likely. Buildings are then identified by examining image objects that fall within the search area defined by the extracted shadows. The image objects examined at this processing stage are generated via segmentation of the PS-MS imagery, not from the DMP. MS segmentation is utilized for generation of the image objects for two reasons. First, only buildings with a bright PAN response are visible in the DMP (see Figs. 5-1, 5-2), and these are extracted using the techniques described in the previous section. Buildings with dark PAN response are not visible in the DMP because they do not have significant contrast from the background (roads,

impervious surface) to be visible in the DMP. Second, MS segmentation provides an independent source of image objects that are complementary to those generated from the DMP. The region merging MS segmentation algorithm described in Section 3.2 is also used here.

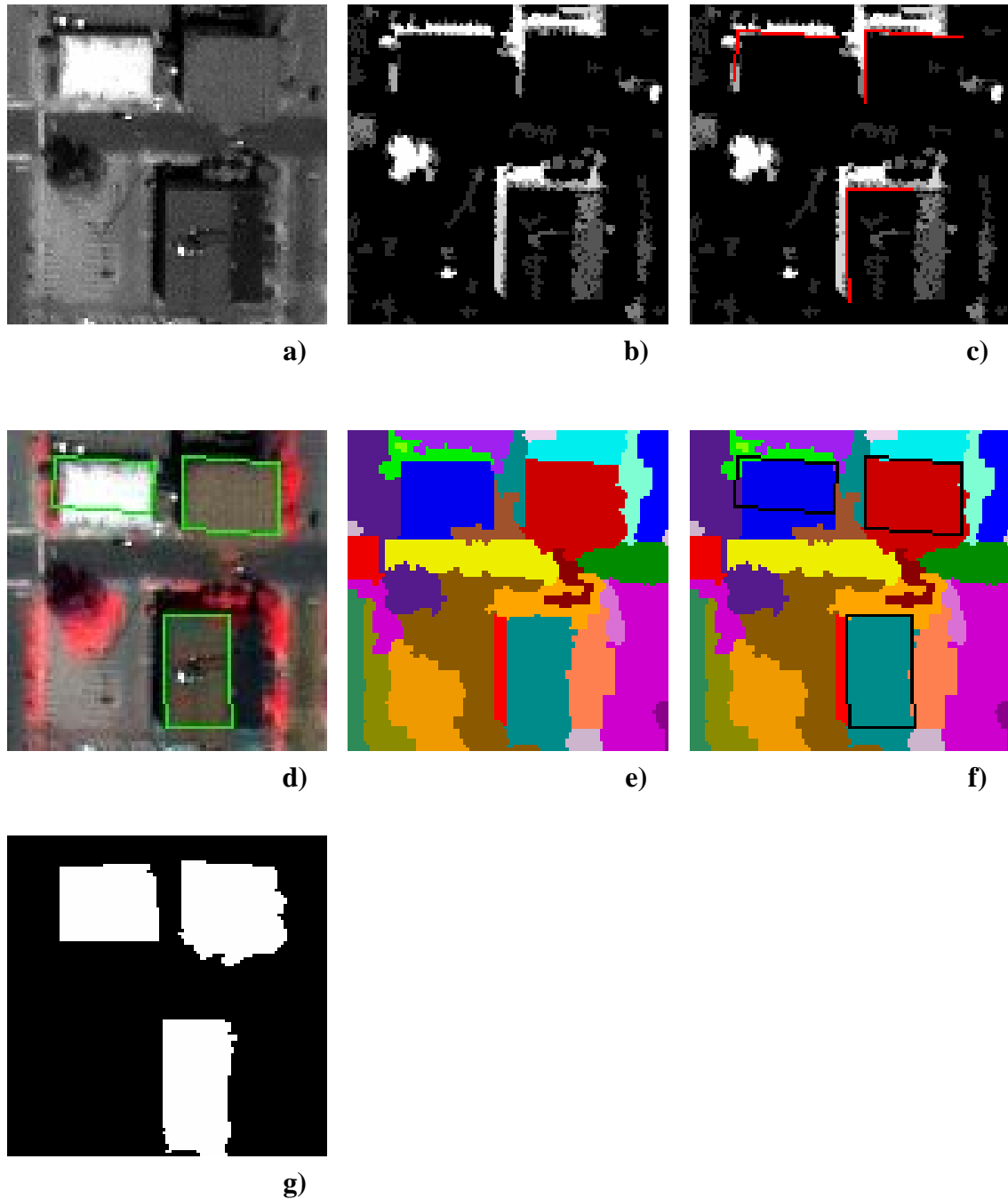
The amount of shadows present in a high-resolution image will vary depending on the sun azimuth and elevation angles and the satellite sensor acquisition azimuth angle. As the sun elevation angle decreases from  $90^{\circ}$ , the length of the shadow buildings cast along the ground grows, and as the sun and sensor azimuth angles get further apart, the amount of shadows visible to the sensor will increase. The IKONOS test image has a sun elevation angle of  $61^{\circ}$ , a sun azimuth angle of  $139^{\circ}$ , and a sensor azimuth angle of  $352^{\circ}$ . As is seen in Fig. 5-2, building shadows are quite prominent in the imagery. Shadows are modeled as consisting of two line segments with approximately  $90^{\circ}$  separation, opening in the direction of the sun azimuth angle. This matches the shape of a shadow cast by the corner of a rectangular building, i.e. the “L-shaped” shadows seen in Fig. 5-2.

The process for identifying buildings from the objects extracted from the closing differential profile is as follows. First, the edges of the closing profile objects in the direction of the sun azimuth angle are extracted with morphological operators. The extracted edge is then approximated by two linear segments. The approximating line segments are constrained such that they must be close to perpendicular and concave in the direction of the sun azimuth angle. A parallelogram is then formed from the piecewise linear edge approximation. This parallelogram, formed from the piecewise linear edge approximation, defines a search area within which there is a high chance of finding a building. All MS segmentation image objects intersecting with the search area

are identified and analyzed to determine if they are buildings. Four shape-based object features are utilized to identify building objects. The MS segmentation image objects are first analyzed using the morphological shape analysis techniques described in Section 3.3-B. The compactness of the object, the percentage overlap between the object and the search area parallelogram, and the percentage of vegetation pixels present in the object, as determined from the NDVI, are also utilized to identify building objects. If none of the MS segmentation objects are identified as buildings, but the dimensions of the parallelogram formed from the shadow edge approximation are similar to that of buildings in the imagery, then the parallelogram is identified as a building, provided it has a low percentage of vegetation pixels. This is done so that if a building has been incorrectly delineated by the MS segmentation, it can still be identified from its shadow. If multiple buildings are detected in the same location, preference is given to buildings identified from the opening differential profile. Figure 5-4 presents a graphical depiction of the process of indirect building detection via shadow identification.

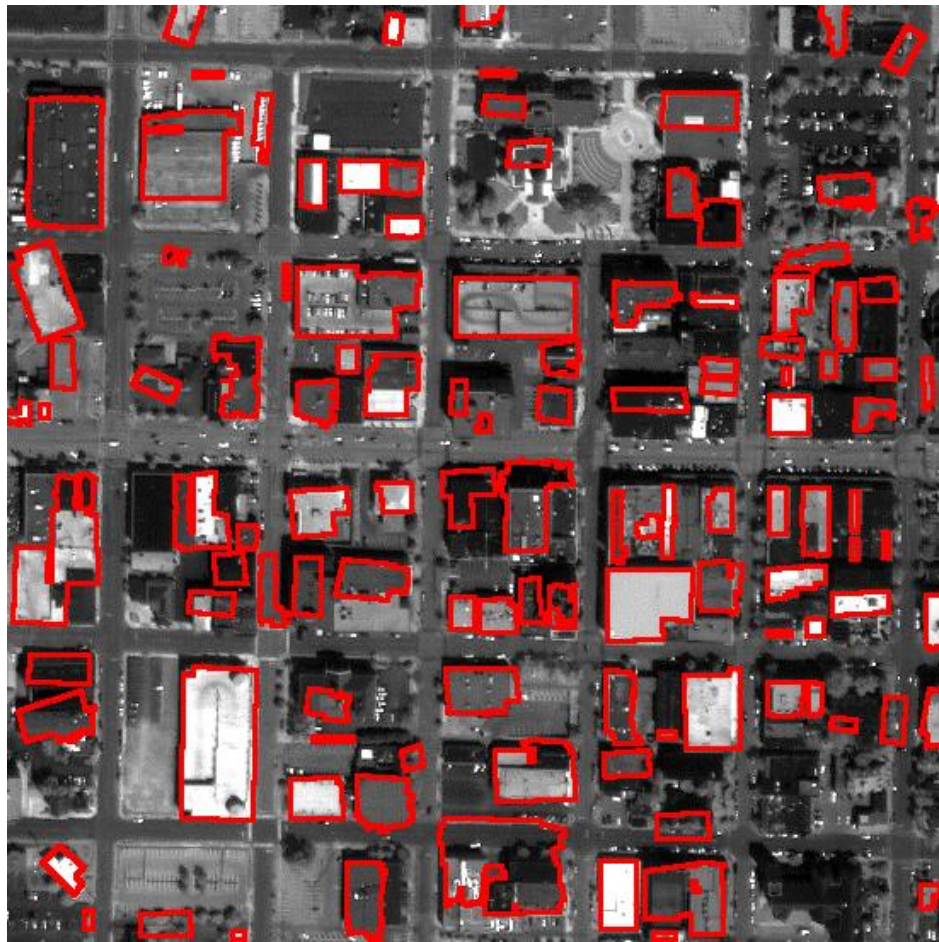
#### **5.4 Test Results**

The building extraction algorithm described in this chapter was tested on a subset of the Columbia IKONOS imagery containing the downtown area. This smaller subset was chosen both because it contained only urban buildings, and due to the length of time it takes to manually generate reference data for an accuracy assessment. Objects in the opening differential profile are identified as buildings if the ratio of the object's area to the approximating polygon is greater than 0.6, the minimum length edge of the bounding rectangle is greater than 5 m, and if the area of the approximating polygon is greater than



**Figure 5-4.** Graphical depiction of indirect building detection via shadow identification. a) PAN imagery, b) closing differential profile ( $r = 5$  m), c) shadow edge approximation, d) parallelogram search areas superimposed over PS-MS imagery, e) segmentation of PS-MS imagery, f) MS segments intersecting with search areas, and g) extracted building objects.

45 m<sup>2</sup>. Objects in the closing differential profile are identified as potential building shadows if the minimum length of the approximating edges is greater than 5 m. MS segmentation objects intersecting the search areas are identified as buildings if they have a fuzzy shape membership greater than 0.5, compactness greater than 0.5, and the percentage of intersection between the object and the search area is at least 10%. Parallelograms formed from shadow edge approximations with no buildings extracted within them are identified as buildings if they have an area of at least 200 m<sup>2</sup>, and if the minimum length edge of the parallelogram is at least 6 m. The results of the building extraction are shown in Fig. 5-5.



**Figure 5-5.** Results of automated building footprint extraction algorithm for Columbia IKONOS PAN image of dense urban area.

The accuracy of the building extraction algorithm is quantified using the accuracy assessment statistics described in Section 4.2. The reference building footprints were manually digitized in the imagery and contain all buildings present in the test site. The accuracy assessment was performed at both the object and pixel levels. For the object level accuracy assessment, all reference building footprints that intersect with objects extracted by the algorithm are counted as being correctly identified. Conversely, all objects extracted by the algorithm that intersect with the reference building footprints are counted as correctly extracted objects. The number of correct objects generated by the extraction algorithm, and the number of correctly extracted reference buildings is not necessarily the same. The pixel level accuracy assessment is performed by simply counting the total number of pixels present after finding the intersection of the output of the extraction algorithm and the reference data. The object and pixel level accuracy assessments represent a best and worst case assessment, respectively, of the quality of the building extraction algorithm. The object level accuracy assessment is displayed in Table 5-1 and the pixel level accuracy assessment is displayed in Table 5-2. The high correctness value of the object-level extraction indicates that most of the extracted building objects are at least partially overlapping with the actual buildings present in the image. However, the lower completeness value indicates that almost one third of the buildings present in the image are not being identified at all. The accuracy of the algorithm is significantly lower when assessed at the pixel level. The low completeness value in the pixel level accuracy assessment is caused by discrepancies between the shape of the buildings extracted by the algorithm and the reference building footprints.



**Table 5-1**  
**Object Level Building Extraction Performance Measures**

	<b>Building Extraction Algorithm</b>	<b>Reference Building Footprints</b>
<b>Correctly Extracted Objects:</b>	97	82
<b>Total Objects:</b>	111	116
	<i>Correctness = 87.4%</i>	<i>Completeness = 70.7%</i>
	<i>Quality = 66.9%</i>	

**Table 5-2**  
**Pixel Level Building Extraction Performance Measures**

	<b>Building Extraction Algorithm</b>	<b>Reference Building Footprints</b>
<b>Correctly Extracted Pixels:</b>	55131	55131
<b>Total Pixels:</b>	71354	110035
	<i>Correctness = 77.3%</i>	<i>Completeness = 50.1%</i>
	<i>Quality = 43.7%</i>	

## 5.5 Summary

In this chapter a fully automated algorithm for 2-D building footprint extraction from high-resolution MS satellite imagery was presented. The algorithm is based on a multi-detector fusion strategy where buildings and their shadows are extracted from the Differential Morphological Profile (DMP) of panchromatic imagery and a segmentation of the pan-sharpened multispectral imagery. The DMP is a multi-scale image analysis technique that provides information about both the size and contrast of multi-scale structures in the image. Buildings with a bright spectral response are extracted from the opening portion of the differential profile, while shadows are extracted from the closing portion of the differential profile. The extracted shadows define a search area where the presence of a building is likely, which is then utilized to identify building objects from

the segmented PS-MS image. The algorithm was tested on a subset of the Columbia IKONOS imagery containing dense urban land cover and was found to produce an accurate (correctness values of 87% and 77% for the object and pixel assessments, respectively) yet somewhat incomplete (completeness values of 71% and 50% for the object and pixel assessments, respectively) building footprint extraction.

The material presented in this chapter was presented at the 2004 *International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2004]. The main contribution of this work to the body of literature on automated building footprint extraction from high-resolution satellite imagery is the development of an approach that makes use of the DMP for the purpose of building extraction. The accuracy of the algorithm compares favorably with results reported for other techniques where completeness values of 72%-81% and quality values of 30-51% were reported [Lee *et al.*, 2003], [Croitoru and Doytsher, 2003].

## Chapter 6

### Fully-Automated Urban Land Cover Classification

#### 6.1 Introduction

In the previous two chapters, automated techniques for road network and building footprint extraction were presented. Both of these techniques are fully automated and require no user interaction beyond the setting of several parameters before the algorithms begin. Pixel-based and object-based fuzzy supervised urban land cover classification approaches were presented in Chapters 2 and 3, respectively. Unlike the fully automated feature extraction techniques, both of these classifiers require human generated training data and are thus only semi-automated. However, if the generation of training data is automated, these supervised classifiers can be utilized in an unsupervised, or self-supervised fashion to perform urban land cover classification. In this chapter, a fully automated technique for classification of urban land cover is presented. Feature extraction techniques are utilized to generate training data that are then input into the supervised classification algorithms, thereby producing a self-supervised urban land cover classifier. These feature extraction techniques do not seek to extract all features present in the imagery. Instead, they are used to identify very high confidence instances of the different urban land cover classes, so as not to input incorrect training data into the classifier. This self-supervised classifier is based on the classification and feature extraction techniques developed in this research, as well as several additional fully

automated techniques for the identification of training data from other classes of interest in the urban environment.

The development of a fully automated urban land cover classification technique would be a significant contribution to the area of satellite image analysis, as there are currently no published results in the literature on this topic. Section 1.1 highlighted many of the civilian, intelligence, and defense applications for which a fully automated classification technique would be very useful. One area where a technique such as this is greatly needed is automated image archival and retrieval. Such a technique would significantly enhance the automated exploitation of the rapidly increasing volumes of satellite image data.

## **6.2 Fully Automated Training Data Generation**

In order to utilize the urban land cover classification algorithms described in Chapters 2 and 3 in an unsupervised or self-supervised fashion, the training data must be generated automatically. Fully automated feature extraction techniques, such as those described in Chapters 4 and 5, are utilized for this purpose. Because our goal here is to generate training data, not produce a complete extraction of the features of interest, the correctness of the extracted features is much more important than the completeness of the features. If incorrectly extracted features are used as training data, the errors will propagate through the classification process and lead to poor classification accuracies. The only concern in terms of the completeness of the extracted features is that a representative sample of the different spectral responses and shapes of the feature classes are obtained from the extraction. The strategy adopted here for the training data

generation for each of the urban land cover classes is to output a fuzzy membership value for each extracted feature, representing a confidence level that the extracted feature is a valid member of the particular land cover class. Using these membership values, features with high confidence can be selected and used as training data for each land cover class.

The same urban land cover classes used for the fuzzy-pixel based classification in Chapter 2 (*Road, Building, Grass, Tree, Bare Soil, Water, and Shadow*) are also used here. To generate training data for the *Road, Building, and Shadow* classes, the automated feature extraction algorithms described in Chapters 4 and 5 have been modified to output confidence values for each of the extracted features. The training data for the vegetation classes is generated by first identifying vegetation areas in the image with the NDVI, followed by texture analysis to discriminate between *Grass* and *Tree* land cover classes. Confidence values are then assigned to the identified vegetation pixels. Training data for the *Water* class is generated through analysis of the DMP and the NDVI. The *Bare Soil* training data is generated through identification of the unique spectral signature of bare soil. Previous results indicate that none of the spatial measures examined thus far have significant effect on the classification accuracy of the *Bare Soil* class (see Section 2.5), so shape or texture based identification of this class is not possible. The training data generation for each of the urban land cover classes is discussed in greater detail in the following subsections.

#### *A. Road Training Data Generation*

The road network extraction algorithm presented in Chapter 4 was modified to output a fuzzy membership value for each extracted road, indicating the confidence level

that the extracted road is valid. The fuzzy membership confidence values are based on the length of the initial line segment detected for each road and the percentage of non-vegetative pixels present in the extracted road, as measured by the NDVI. Roads consisting of long line segments and low percentages of vegetation receive high confidence values of being correct. The membership values for large length and low percentage of non-vegetated pixels are computed using *S*-functions (see Section 3.3-A) as follows:

$$f_L(L) = S(L ; 0 , 80 , 300) \quad (6-1)$$

$$f_{nv}(p_{nv}) = S(p_{nv} ; 0.5 , 0.75 , 1.0), \quad (6-2)$$

where  $L$  is the length of the initially identified line segment and  $p_{nv}$  is the percentage of non-vegetated pixels present in the extracted road. These values are combined using the standard fuzzy intersection operator,

$$f_{road} = \min[f_L , f_{nv}]. \quad (6-3)$$

$f_{road}$  is the confidence that the extracted road is valid and is used to identify high confidence extracted roads for use as training data.

### *B. Building and Shadow Training Data Generation*

The fully automated 2-D building footprint extraction algorithm presented in Chapter 5 has also been modified to output confidence values for the extracted building footprints and shadows. The crisp decision thresholds for building identification have been fuzzified to compute a membership value for each potential building and shadow. Membership values for buildings detected in the opening differential profile are computed using *S*-functions as follows:

$$f_r(r_A) = S(r_A ; 0.2 , 0.6 , 1.0) \quad (6-4)$$

$$f_A(A_p) = S(A_p ; 30 , 45 , 60) \quad (6-5)$$

$$f_{min}(L_{min}) = S(L_{min} ; 2 , 5 , 8) \quad (6-6)$$

$$f_{max}(L_{max}) = 1 - S(L_{max} ; 125 , 150 , 175) \quad (6-7)$$

where  $r_A$  is the ratio between the object's area and the area of the approximating polygon,  $A_p$  is the area of the approximating polygon, and  $L_{min}$  and  $L_{max}$  are the magnitude of the minimum and maximum length sides of the minimum area bounding rectangle of the object, respectively. The fuzzy membership values are combined using the standard fuzzy intersection operator to calculate

$$f^l_{building} = \min[f_r, f_A, f_{min}, f_{max}], \quad (6-8)$$

which is the confidence of the building object extracted from the opening differential profile. Objects detected in the opening profile receive high membership values if the ratio between the area of the object and the approximating polygon is near 1.0, and the dimensions of the minimum area bounding rectangle are greater than about 8 m and less than about 125 m.

As described in Section 5.3-B, there are two methods for extracting building objects indirectly through identification of their cast shadows. First, all MS segments within the search area parallelogram defined by the shadow edge are examined. If no MS segments meeting the building criteria are identified and the search area parallelogram has the dimensions of a building, then the parallelogram is identified as a building object. The process of identifying MS segments within the search area is fuzzified as follows:

$$f_{SA} = S(P_{SA} ; 0.0 , 0.1 , 0.3) \quad (6-9)$$

$$f_C = S(C ; 0.2 , 0.5 , 0.8) \quad (6-10)$$

$$f^2_{building} = \min[f_{SA}, f_C, f_{rect}] \quad (6-11)$$

where  $P_{SA}$  is the percentage of the MS segment within the search area,  $C$  is the compactness of the MS segment, and  $f_{rect}$  is the fuzzy shape membership of the segment being ‘approximately rectangular’, as defined in Section 3.3-B. MS segmentation objects receive high membership values if greater than about 30% of the object is within the search area, the compactness is high, and the fuzzy shape membership of being ‘approximately rectangular’ is high.

The process of identifying the search area parallelogram as a building object in absence of an MS segment building object is fuzzified as follows:

$$f_A = S(A_{SA}; 100, 200, 300) \quad (6-12)$$

$$f_{min} = S(L_{min}; 3, 7, 9) \quad (6-13)$$

$$f_{total} = S(L_{total}; 15, 25, 35) \quad (6-14)$$

$$f^3_{building} = \max[f_{total}, \min[f_A, f_{min}]] \quad (6-15)$$

where  $A_{SA}$  is the area of the search area parallelogram,  $L_{min}$  is the minimum length of the shadow edge, and  $f_{total}$  is the sum of the lengths of the shadow edge. Expressed logically, Eq. 6-15 is  $f_{total}$  OR  $f_A$  AND  $f_{min}$ . The search area parallelogram receives a high confidence value if the area of the parallelogram is greater than about 300 m<sup>2</sup>, and the minimum length of the shadow edge is greater than about 9 m, or the total length of the shadow edge is greater than about 35 m. Because three techniques for building identification are utilized, it is possible for multiple building objects to be identified in a single location. If multiple overlapping objects are detected, the object with the largest membership value is retained.



The fuzzy confidence values for the extracted shadows, from which the search area parallelograms are formed, are calculated using Eqs. 6-12 through 6-15 with an additional requirement that the average PAN response of the shadow object be spectrally dark:

$$f_{PAN}(PAN) = 1 - S(PAN ; 285 , 350 , 400) \quad (6-16)$$

$$f_{shadow} = \min[f_{building}^3, f_{PAN}] \quad (6-17)$$

Extracted shadow edges receive high confidence values if they have an average PAN response less than about 285 and the search area parallelograms formed from them have high membership.

### C. Grass and Tree Training Data Generation

It was demonstrated in Chapter 2 that the entropy texture measure is able to increase discrimination between the spectrally similar *Grass* and *Tree* classes. Here, entropy is used in conjunction with the NDVI to identify training data for the vegetative classes. First, the NDVI is used to identify pixels with vegetative land cover. Then, the entropy texture measure is utilized to identify high confidence instances of *Grass* and *Tree* from within the pixels identified as containing vegetative land cover. The first order entropy texture measure is calculated using an 11x11 pixel window. The fuzzy membership values are calculated as follows:

$$f_{veg} = S(NDVI ; 0.2 , 0.26 , 0.35) \quad (6-18)$$

$$f_{high\ ent} = S(Entropy ; 3.9 , 4.1 , 4.3) \quad (6-19)$$

$$f_{low\ ent} = 1 - S(Entropy ; 3.6 , 3.9 , 4.2) \quad (6-20)$$

$$f_{tree} = \min[f_{veg}, f_{high\ ent}] \quad (6-21)$$

$$f_{grass} = \min[f_{veg}, f_{low\ ent}]. \quad (6-22)$$

Pixels with both high NDVI values and high entropy values receive large membership values in the *Tree* class, whereas pixels with high NDVI values and low entropy values receive large membership values in the *Grass* class.

#### D. Bare Soil Training Data Generation

As was demonstrated in Chapter 2, it is possible to accurately identify the *Bare Soil* class utilizing only its spectral characteristics. It was also shown that none of the spatial measures investigated improved the ability to identify this class. Several additional spatial measures have been investigated, including the DMP and additional texture measures, but no spatial measures have been found with the ability to accurately identify *Bare Soil* pixels independent of the spectral signature. The difficulty of utilizing spatial features for the identification of this class is that the spatial characteristics of bare soil vary widely, with no regularity. Even within a single image, the size, shape, and amount of texture present in this land cover class can vary significantly. It is therefore not possible to produce a spatial model for bare soil. To generate training data for the *Bare Soil* class, the unique spectral signature of this land cover class is utilized to form fuzzy membership functions as follows:

$$f^1_N(NIR) = \pi(NIR ; \mu^1_N - 3\sigma, \mu^1_N - 2\sigma, \mu^1_N - \sigma, \mu^1_N + \sigma, \mu^1_N + 2\sigma, \mu^1_N + 3\sigma) \quad (6-23)$$

$$f^1_R(RED) = \pi(RED ; \mu^1_R - 3\sigma, \mu^1_R - 2\sigma, \mu^1_R - \sigma, \mu^1_R + \sigma, \mu^1_R + 2\sigma, \mu^1_R + 3\sigma) \quad (6-24)$$

$$f^1_G(GREEN) = \pi(GREEN ; \mu^1_G - 3\sigma, \mu^1_G - 2\sigma, \mu^1_G - \sigma, \mu^1_G + \sigma, \mu^1_G + 2\sigma, \mu^1_G + 3\sigma) \quad (6-25)$$

$$f^1_B(BLUE) = \pi(BLUE ; \mu^1_B - 3\sigma, \mu^1_B - 2\sigma, \mu^1_B - \sigma, \mu^1_B + \sigma, \mu^1_B + 2\sigma, \mu^1_B + 3\sigma) \quad (6-26)$$

$$f^1_{bare\ soil} = \min[f^1_N, f^1_R, f^1_G, f^1_B] \quad (6-27)$$

$$f^2_N(NIR) = \pi(NIR ; \mu^2_N - 3\sigma, \mu^2_N - 2\sigma, \mu^2_N - \sigma, \mu^2_N + \sigma, \mu^2_N + 2\sigma, \mu^2_N + 3\sigma) \quad (6-28)$$

$$f^2_R(RED) = \pi(RED ; \mu^2_R - 3\sigma, \mu^2_R - 2\sigma, \mu^2_R - \sigma, \mu^2_R + \sigma, \mu^2_R + 2\sigma, \mu^2_R + 3\sigma) \quad (6-29)$$

$$f^2_G(GREEN) = \pi(GREEN ; \mu^2_G - 3\sigma, \mu^2_G - 2\sigma, \mu^2_G - \sigma, \mu^2_G + \sigma, \mu^2_G + 2\sigma, \mu^2_G + 3\sigma) \quad (6-30)$$

$$f^2_B(BLUE) = \pi(BLUE ; \mu^2_B - 3\sigma, \mu^2_B - 2\sigma, \mu^2_B - \sigma, \mu^2_B + \sigma, \mu^2_B + 2\sigma, \mu^2_B + 3\sigma) \quad (6-31)$$

$$f^2_{bare\ soil} = \min[f^2_N, f^2_R, f^2_G, f^2_B] \quad (6-32)$$

$$f_{bare\ soil} = \max[f^1_{bare\ soil}, f^2_{bare\ soil}], \quad (6-33)$$

where  $\mu^1_N = 500$ ,  $\mu^1_R = 350$ ,  $\mu^1_G = 380$ ,  $\mu^1_B = 320$ ,  $\mu^2_N = 420$ ,  $\mu^2_R = 360$ ,  $\mu^2_G = 370$ ,  $\mu^2_B = 300$ , and  $\sigma = 25$ . Two membership functions are utilized,  $f^1_{bare\ soil}$  and  $f^2_{bare\ soil}$ , because it was found that two distinct spectral signatures of *Bare Soil* were present in the imagery. These values were determined through examination of several sites in the imagery containing bare soil land cover.

#### E. Water Training Data Generation

While the appearance of different bodies of water such as rivers, streams, lakes, and ponds varies significantly in high-resolution imagery, if one of these types of water bodies can be automatically identified, the extracted pixels can be used to train the classifier. Of the above-mentioned water body types, small lakes and ponds have the least variability in appearance, typically appearing as large, dark compact objects with no vegetation present. Objects fitting this profile are easily identified in the closing differential profile. Because only large objects are of interest, the closing differential profile computed using a structuring element with radius  $r = 26$  m is selected for analysis. This approach presupposes the existence of water bodies at this scale. However, this limitation can be avoided by using a very large radius SE to compute the DMP at this level, thus identifying all large dark objects in the imagery. Confidence values are

computed based on the strength of the DMP response, the amount of vegetation present in the object, and the area of the object as follows:

$$f_{DMP}(CL_{26}) = S(CL_{26} ; 5 , 6 , 7) \quad (6-34)$$

$$f_{veg}(P_{veg}) = 1 - S(P_{veg} ; 0.20 , 0.35 , 0.45) \quad (6-35)$$

$$f_{area}(A) = S(A ; 1000 , 2000 , 3000) \quad (6-36)$$

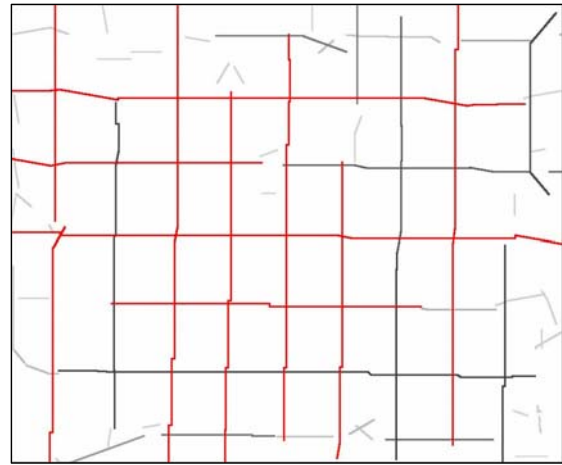
where  $CL_{26}$  is the response of closing differential profile level computed with a structuring element of  $r = 26$  m,  $P_{veg}$  is the percentage of vegetation present in the object, and  $A$  is the area of the object.

#### F. Training Data Generation

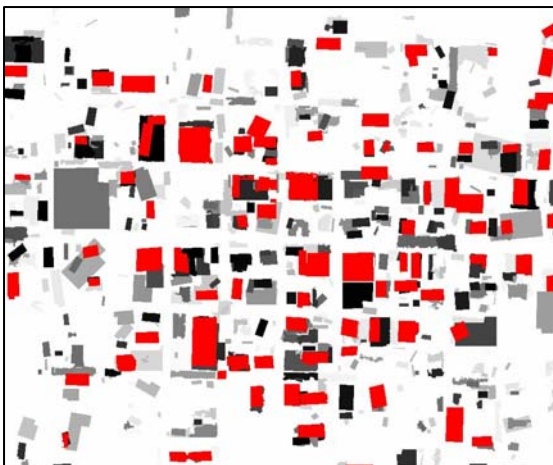
The fully automated feature extraction algorithms described above output a fuzzy confidence value for each extracted feature. Extracted features with high confidence values, indicating valid features, are utilized as training data, and the rest of the extracted features are discarded. This is accomplished by thresholding the fuzzy confidence value of each extracted feature. A value of 0.95 is used to threshold the *Building*, *Shadow*, *Grass*, *Tree*, and *Water* classes. A value of 0.85 is used to threshold the *Road* class, and 0.99 is used for the *Bare Soil* class. These values have been chosen because they produce training data that is both accurate and representative of the variability within each land cover class. The training data generated for the *Road* and *Building* classes in a dense urban area, as well as the fuzzy confidence values for the extracted features, are shown in Fig. 6-1. The training data generated for the *Grass* and *Tree* classes for an area with suburban land cover, as well as the fuzzy confidence values for the extracted features, are shown in Fig. 6-2.



a)



b)



c)

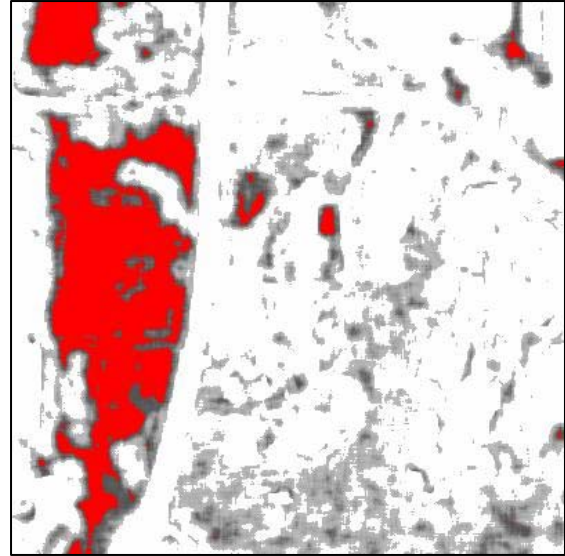
**Figure 6-1.** Automatically generated training data from dense urban area for b) *Road* class, and c) *Building* class. Training data shown in red, high confidence features shown in dark gray, and low confidence features shown in light gray.

### 6.3 Self-Supervised Urban Land Cover Classification

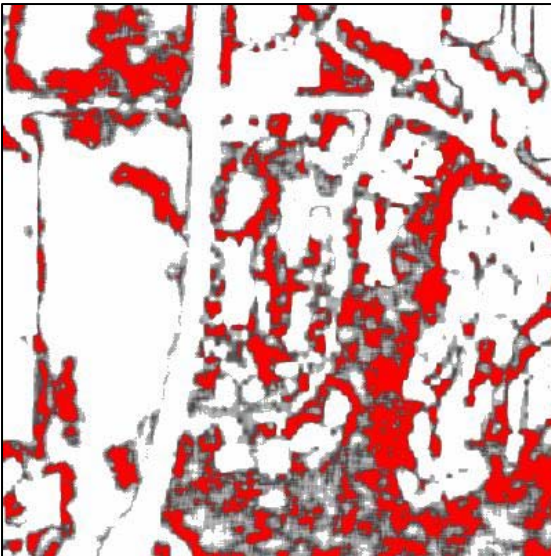
Once high confidence instances of each of the urban land cover classes have been identified, this data can then be utilized to train a supervised classification system. Because labeled training data is generated internally by the system, systems of this type can be referred to as self-supervised. Self-supervised classification systems differ from unsupervised classifiers in that unsupervised classifiers output an unlabeled classification, requiring further analysis to determine the class labels, whereas self-supervised classifiers output a labeled classification. The supervised classification scheme utilized here follows



a)



b)



c)

**Figure 6-2.** Automatically generated training data from residential area for b) *Grass* class, and c) *Tree* class. Training data shown in red, high confidence features shown in dark gray, and low confidence features shown in light gray.

that described in Chapters 2 and 3. The automatically generated training data is first used to train a maximum likelihood classifier, followed by the hierarchical pixel-based fuzzy classifier, and finally the object-based fuzzy logic classifier.

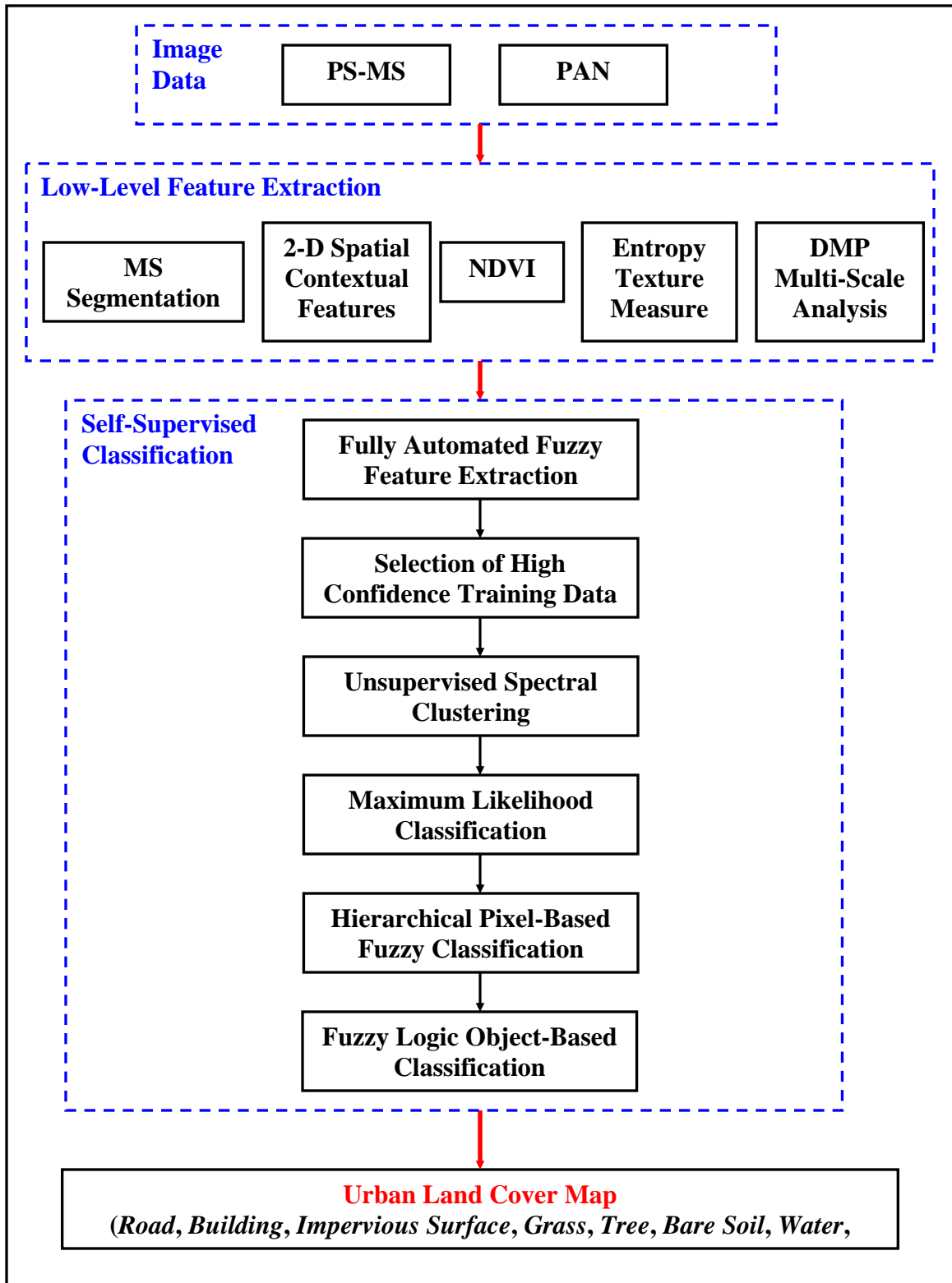
Because of spectral variation within individual land cover classes, it is necessary to train the supervised classifier on multiple sub-classes within each urban land cover

class (see Section 2.2). Following classification, the sub-classes from each land cover class are combined. To accommodate within class spectral variation, the automatically generated training data from each class is divided into sub-classes via unsupervised clustering before maximum likelihood classification is performed. After maximum likelihood classification, the sub-classes are combined into the urban land cover classes of interest and the classification process continues through the fuzzy pixel-based and object-based approaches, as described in Chapters 2 and 3. A flow chart summarizing the self-supervised classification approach is presented in Fig. 6-3.

Unsupervised clustering is performed utilizing the standard  $k$ -means clustering algorithm to divide the extracted training data into spectrally coherent sub-classes. The *Road* and *Building* classes are each divided into 10 sub-classes; the *Grass* and *Tree* classes are each divided into 5 sub-classes; the *Bare Soil* class is divided into 2 sub-classes; the *Water* and *Shadow* classes are not subdivided.

#### **6.4 Test Results**

The fully automated self-supervised classification scheme was applied to a subset of the IKONOS image of Columbia, Missouri (Fig. 2-1). The test site is shown in Fig. 3-4 and consists primarily of dense urban land cover. An accuracy assessment of the resulting classifications was performed making use of the same reference pixel datasets utilized in Chapter 3 to assess the accuracy of the supervised fuzzy object-based classifier. See Section 3.4 for a detailed description of the reference pixel datasets. Accuracy assessments were performed for each of the classification outputs (maximum likelihood, fuzzy pixel, and fuzzy object) produced by the self-supervised classification scheme. The



**Figure 6-3.** Processing flow chart summarizing the steps of the self-supervised classification scheme.



individual class accuracies and the overall classification accuracy for each output, as well as the corresponding results from the manually supervised classifiers, are displayed in Tables 6-1 through 6-3. The confusion matrices for the self-supervised classification outputs are displayed in Tables 6-4 through 6-6. For comparison purposes, the confusion matrices for the manually supervised classifications are displayed in Tables 6-7 through 6-9.

**Table 6-1**  
**Accuracies of Maximum Likelihood Classifications**

	<b>Supervised</b> <b>(%)</b>	<b>Self-Supervised</b> <b>(%)</b>
<i>Road</i>	84.4	80.2
<i>Building</i>	83.1	63.7
<i>Grass</i>	92.8	93.8
<i>Tree</i>	83.5	91.9
<b><i>Overall Accuracy</i></b>	<b>85.9</b>	<b>82.4</b>

**Table 6-2**  
**Accuracies for Fuzzy Pixel-Based Classifications**

	<b>Supervised</b> <b>(%)</b>	<b>Self-Supervised</b> <b>(%)</b>
<i>Road</i>	97.9	97.2
<i>Building</i>	90.7	93.2
<i>Grass</i>	94.5	100.0
<i>Tree</i>	96.2	79.6
<b><i>Overall Accuracy</i></b>	<b>94.8</b>	<b>92.5</b>

**Table 6-3**  
**Accuracies for Fuzzy Object-Based Classifications**

	<b>Supervised</b>	<b>Self-Supervised</b>
	<b>(%)</b>	<b>(%)</b>
<i>Road</i>	99.2	95.0
<i>Building</i>	76.1	70.1
<i>Imp. Surf.</i>	81.0	72.2
<i>Grass</i>	91.3	100.0
<i>Tree</i>	99.9	99.1
<b><i>Overall Accuracy</i></b>	<b>89.5</b>	<b>87.4</b>

For each of the classifier outputs, the overall accuracy of the self-supervised classification is only 2-3% lower than that of the manually supervised classification. There is a significant decrease in the *Building* class accuracy of approximately 20% between the self-supervised and manually supervised maximum likelihood classification, as can be seen in Table 6-1. This is due to overclassification of the *Road* class. However, the problem is solved by the hierarchical fuzzy pixel-based classification stage, where the average accuracies of the *Road* and *Building* classes exceed that of the manually supervised classification. There is a 17% decrease in the *Tree* class between the manual and self-supervised fuzzy pixel-based classification due to over classification of the *Grass* class. This error is unexpected and believed to be due to the fact that all of the automatically generated training data for the *Tree* class is extracted from highly textured areas. While appropriate for the majority of this class, there are areas within the *Tree* land cover class that are not highly textured, possibly due to trees with very large and homogeneous crowns. Because the texture of these instances of *Tree* land cover matches that of *Grass*, they are misclassified. The problem of misclassification of *Tree* reference pixels as *Grass* is solved in the object-based stage of the self-supervised

classification. As seen in Table 6-3, the self-supervised classifier produces a classification with virtually no errors in the reference data of the vegetative classes. The object-based classifier is able to correct this problem because the proportions of each class present in the object are used as features, resulting in a majority filtering type operation within the object. The regions where *Tree* land cover are incorrectly identified as *Grass* in the self-supervised pixel-based classification are all quite small and are removed by the majority filtering effect of the object-based classifier.

**Table 6-4**  
**Confusion Matrix for Self-Supervised Maximum Likelihood Classification**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	Total	%
<i>Road</i>	8227	3601	0	0	11828	84.4
<i>Building</i>	2025	6563	0	7	8596	82.1
<i>Grass</i>	0	0	9588	827	10415	86.7
<i>Tree</i>	0	0	629	9458	10087	92.2
<i>Shadow</i>	0	135	0	0	135	-
<b>Total</b>	10252	10257	10217	10292	41060	
<b>%</b>	80.3	63.7	93.8	91.9		82.4

**Table 6-5**  
**Confusion Matrix for Self-Supervised Pixel-Based Fuzzy Classification**

	<i>Road</i> Ref.	<i>Building</i> Ref.	<i>Grass</i> Ref.	<i>Tree</i> Ref.	Total	%
<i>Road</i>	9974	564	0	0	10538	94.7
<i>Building</i>	278	9600	0	7	9885	97.1
<i>Grass</i>	0	0	10217	2095	12312	83.0
<i>Tree</i>	0	0	0	8190	8190	100.0
<i>Shadow</i>	0	135	0	0	135	-
<b>Total</b>	10252	10299	10217	10292	41060	
<b>%</b>	97.3	93.2	100.0	79.6		92.5

**Table 6-6**  
**Confusion Matrix for Self-Supervised Object-Based Fuzzy Classification**

	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Imp. Surf. Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	9737	499	412	0	0	10648	91.4
<b>Building</b>	9	7290	2457	0	0	9756	74.7
<b>Imp. Surf.</b>	475	2487	7437	0	0	10399	71.5
<b>Grass</b>	0	0	0	10217	89	10306	99.1
<b>Tree</b>	31	0	0	0	10203	10234	99.7
<b>Shadow</b>	0	27	0	0	0	27	-
<b>Total</b>	10252	10303	10306	10217	10292	51370	
<b>%</b>	95.0	70.8	72.2	100.0	99.1		87.4

**Table 6-7**  
**Confusion Matrix for Supervised Maximum Likelihood Classification**

	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	8646	1594	0	0	10240	84.4
<b>Building</b>	1577	8523	0	279	10379	82.1
<b>Grass</b>	19	1	9378	1420	10818	86.7
<b>Tree</b>	0	0	723	8593	9316	92.2
<b>Shadow</b>	4	139	0	0	143	-
<b>Total</b>	10246	10257	10101	10292	40896	
<b>%</b>	84.4	83.1	92.8	83.5		85.9

**Table 6-8**  
**Confusion Matrix for Supervised Pixel-Based Fuzzy Classification**

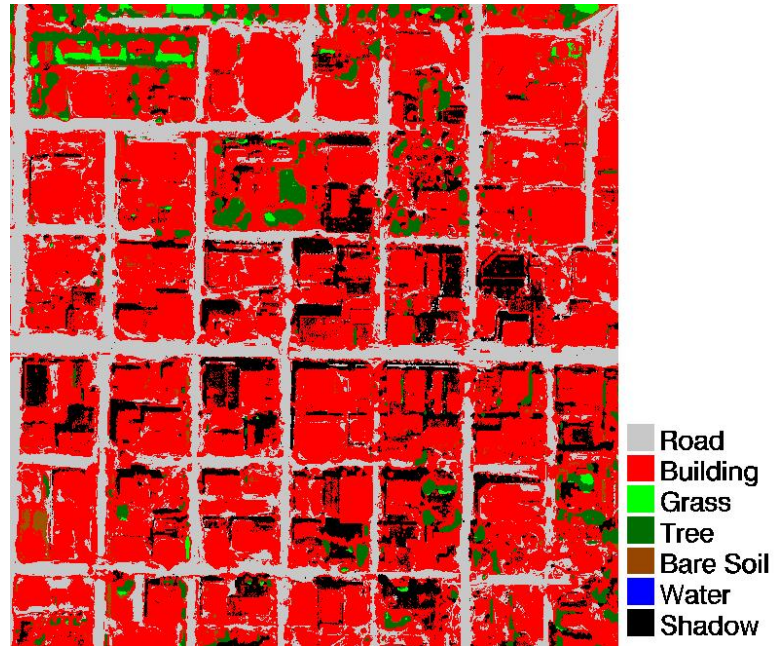
	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	10029	808	0	31	10868	92.3
<b>Building</b>	194	9309	0	248	9751	95.5
<b>Grass</b>	19	1	9540	112	9671	98.7
<b>Tree</b>	0	0	561	9901	10462	94.6
<b>Shadow</b>	1	146	0	0	147	-
<b>Total</b>	10243	10264	10101	10292	40900	
<b>%</b>	97.9	90.7	94.5	96.2		94.8

**Table 6-9**  
**Confusion Matrix for Supervised Object-Based Fuzzy Classification**

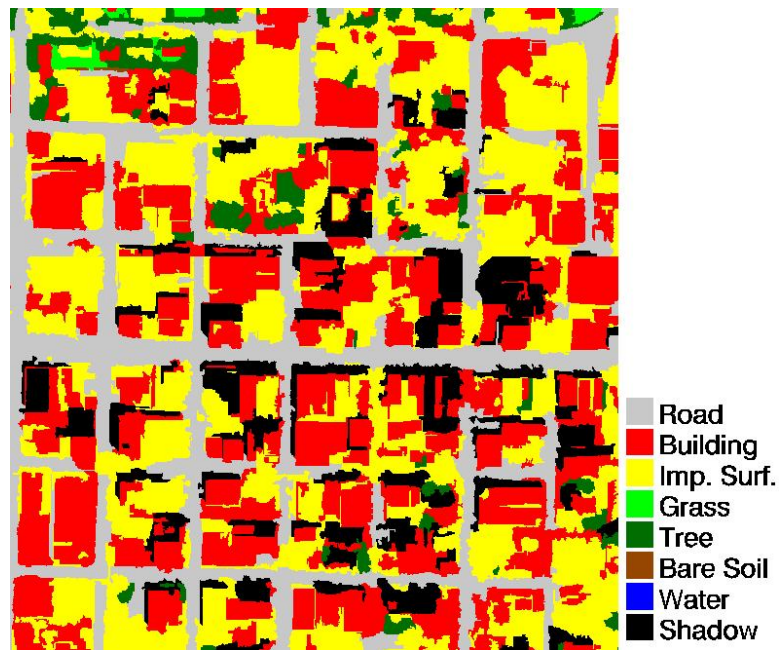
	<b>Road Ref.</b>	<b>Building Ref.</b>	<b>Imp. Surf. Ref.</b>	<b>Grass Ref.</b>	<b>Tree Ref.</b>	<b>Total</b>	<b>%</b>
<b>Road</b>	10164	499	826	0	0	11489	88.5
<b>Building</b>	0	7845	1137	0	0	8982	87.3
<b>Imp. Surf.</b>	86	1959	8343	0	13	10401	80.2
<b>Grass</b>	0	0	0	9330	0	9330	100.0
<b>Tree</b>	0	0	0	887	10279	11166	92.1
<b>Shadow</b>	0	0	0	0	0	0	-
<b>Total</b>	10250	10303	10306	10217	10292	51368	
<b>%</b>	99.2	76.1	81.0	91.3	99.9		89.5

There are 4%, 6%, and 9% decreases in the accuracies of the *Road*, *Building*, and *Impervious Surface* class accuracies, respectively, between the manual and self-supervised object-based classifiers. It is believed that the decrease in the accuracy of the *Building* and *Impervious Surface* classes is partially due to errors in the classification of the *Shadow* class, which is used in the identification of the *Building* class. The self-supervised fuzzy pixel-based and object-based classifications of a dense urban land cover subset image (Fig. 3-1) are shown in Figs. 6-4 and 6-5, respectively. The land cover classification maps of the entire urban test site (Fig. 3-4) for the self-supervised fuzzy pixel-based and object-based classifiers are shown in Figs. 6-6 and 6-7, respectively.

As was discussed in Chapter 3, the object-based classification scheme was originally developed to discriminate between buildings and non-road impervious surface in areas of dense urban land cover, and, accordingly, is not suitable for classification of areas with suburban land cover. To address this issue, the *Building* class was subdivided into *Urban Building* and *Suburban Building*. Training data for the *Suburban Building* class is automatically generated by simply identifying small, dark, non-vegetative objects in the closing differential profile. The *Suburban Building* class is then identified in the object-based classifier as image objects labeled as *Impervious Surface* with dimensions



**Figure 6-4.** Self-supervised hierarchical pixel-based fuzzy classification of dense urban area shown in Fig. 3-1.



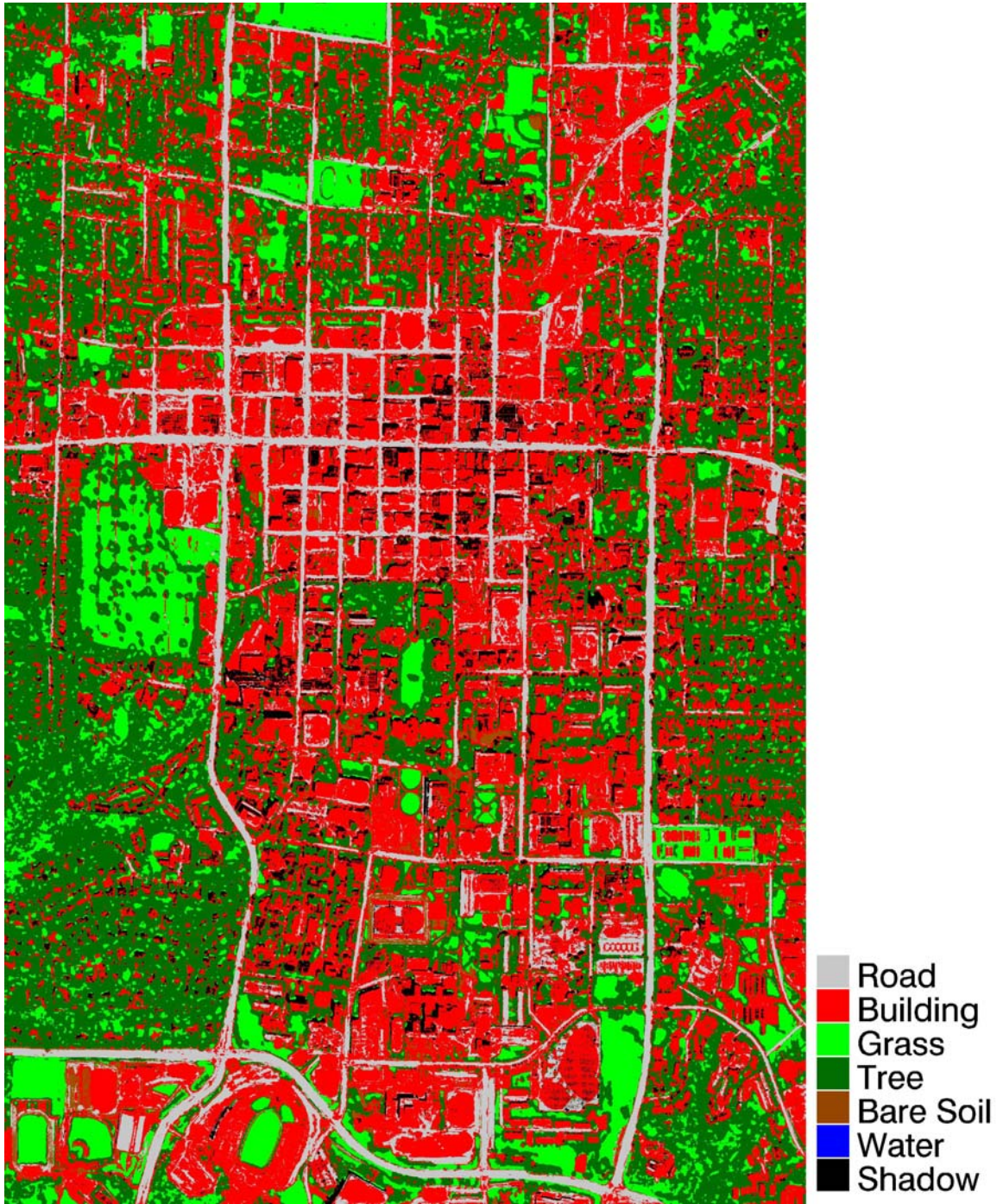
**Figure 6-5.** Self-supervised object-based fuzzy logic classification of dense urban area shown in Fig. 3-1.

between 5 m and 40 m, surrounded by land cover comprised of at least 15% vegetation. The self-supervised land cover classification map of the urban test site (Fig. 3-4) with discrimination between the two building classes is shown in Fig. 6-8, and the object-based self supervised classification of a test site with suburban land cover (Fig. 3-9) is shown in Fig. 6-9.

## **6.5 Summary**

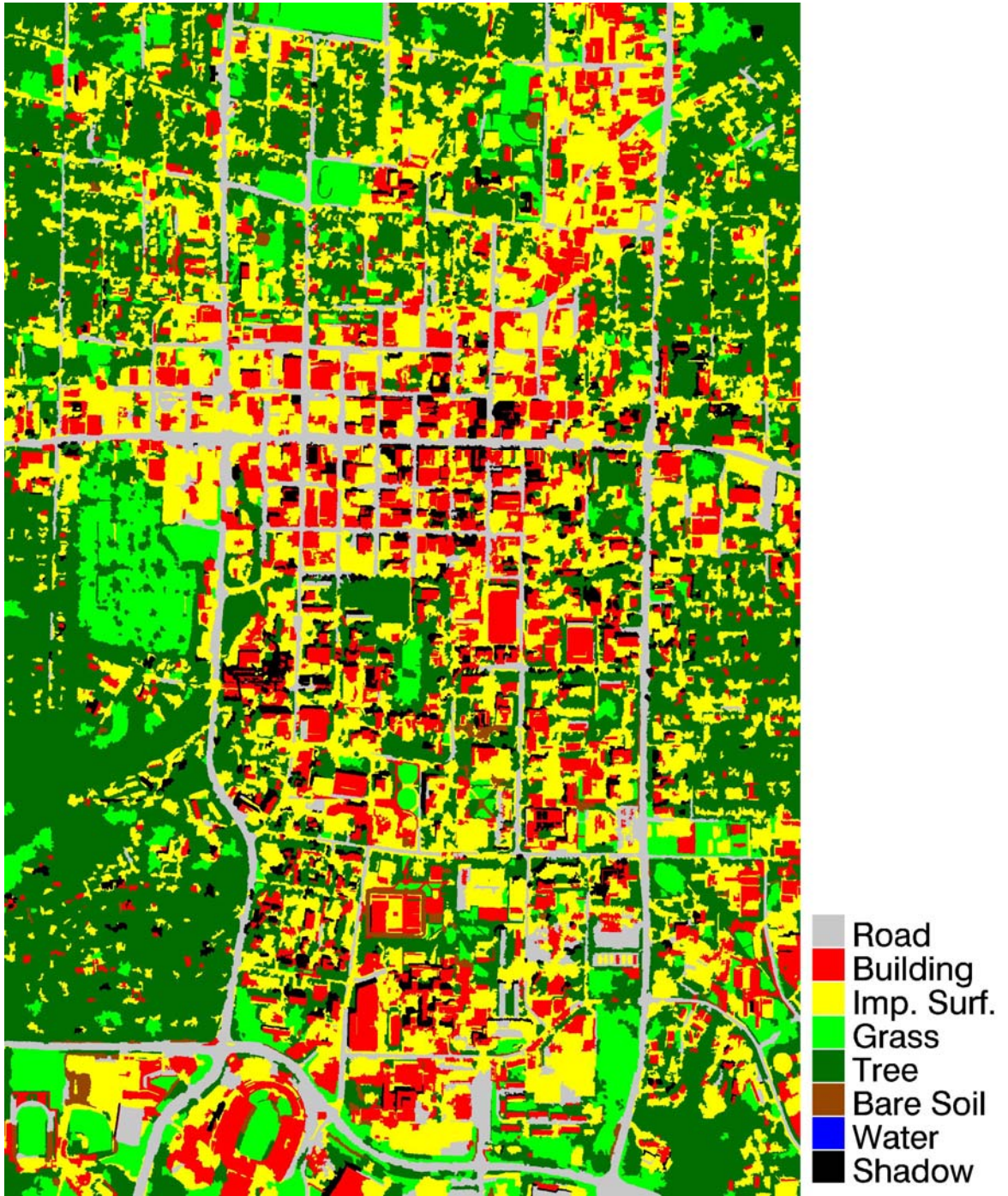
A fully automated self-supervised classification approach for urban land cover classification of high-resolution MS satellite imagery is presented in this chapter. The classifier is based on the supervised classification approaches presented in Chapters 2 and 3, however the training data is automatically generated using feature extraction techniques that identify high confidence instances of the urban land cover features. The automated road network and 2-D building footprint extraction algorithms presented in Chapters 4 and 5 have been modified to output a fuzzy confidence value for each extracted feature. Other spatially and spectrally based feature extraction algorithms have been developed to identify training data for the other urban land cover classes. After feature extraction and selection of high accuracy training data, the extracted features are subdivided into spectrally coherent sub-classes by unsupervised spectral clustering. The training data are then used to train a maximum likelihood classifier, followed by the hierarchical fuzzy pixel-based classifier, and finally the object-based classifier. Test results indicate that the self-supervised classification approach is able to produce urban land cover maps with overall accuracies that are only 2-3% less than that of the manually trained supervised classifiers.





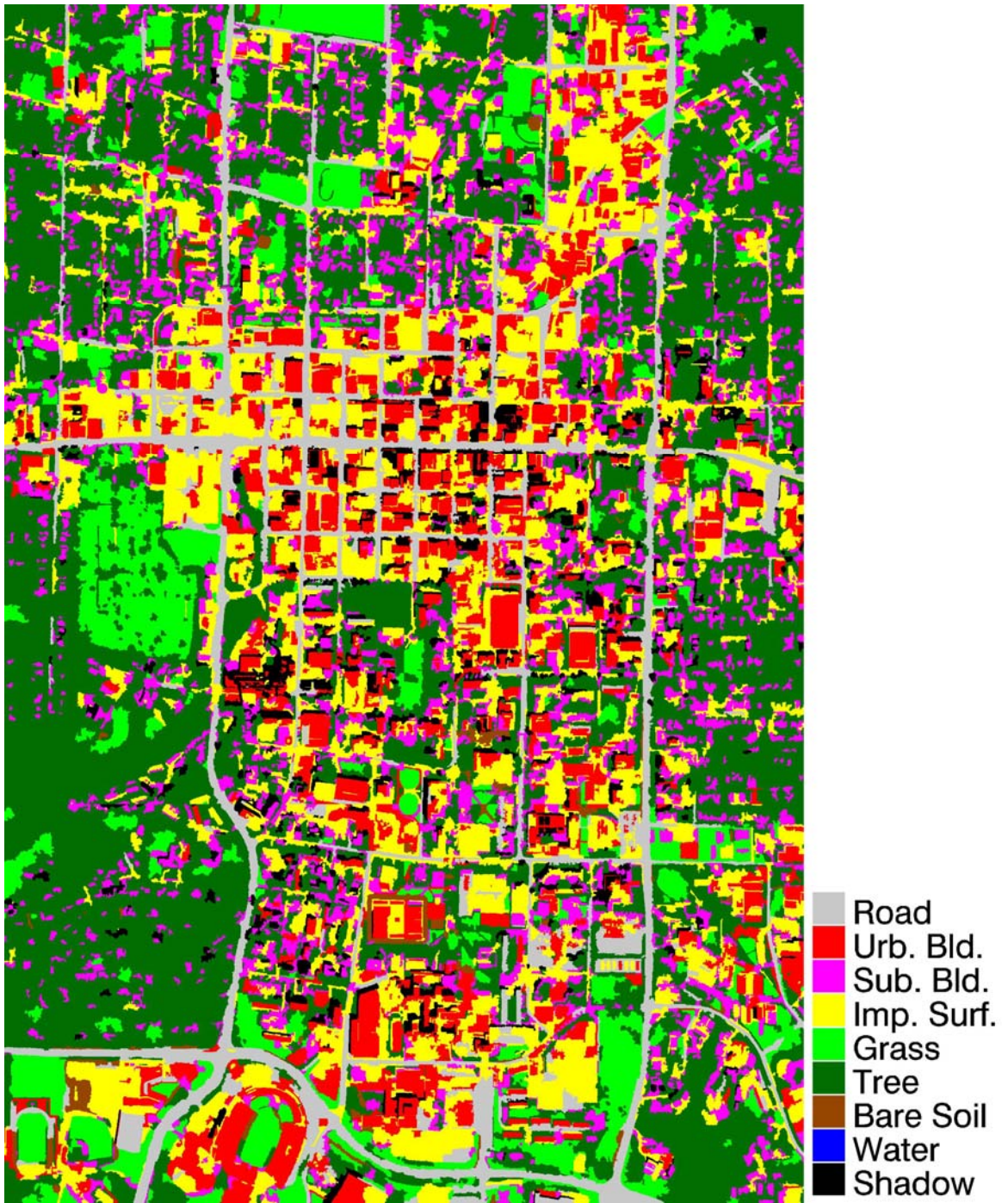
**Figure 6-6.** Self-supervised hierarchical pixel-based fuzzy classification of urban area shown in Fig. 3-4.





**Figure 6-7.** Self-supervised fuzzy logic object-based classification of urban area shown in Fig. 3-4.





**Figure 6-8.** Self-supervised fuzzy logic object-based classification of urban area shown in Fig. 3-4 with differentiation between urban and suburban buildings.





**Figure 6-7.** Self-supervised fuzzy logic object-based classification of suburban area shown in Fig. 3-9 with differentiation between urban and suburban buildings.

The material presented in this chapter will be submitted to the *IEEE Transactions on Geoscience and Remote Sensing* for publication. The major contribution of this work is the development of a fully automated self-supervised classification approach for high-resolution imagery of urban areas. The development of a fully automated urban land cover classification technique is a significant contribution to the area of satellite image analysis, as there are currently no published results in the literature on this topic. Section 1.1 highlighted many of the civilian, intelligence, and defense applications for which a fully automated classification technique would be very useful. One area where a technique such as this is greatly needed is automated image archival and retrieval. A fully automated urban land cover classification technique can significantly enhance the exploitation of the rapidly increasing volumes of satellite image data.

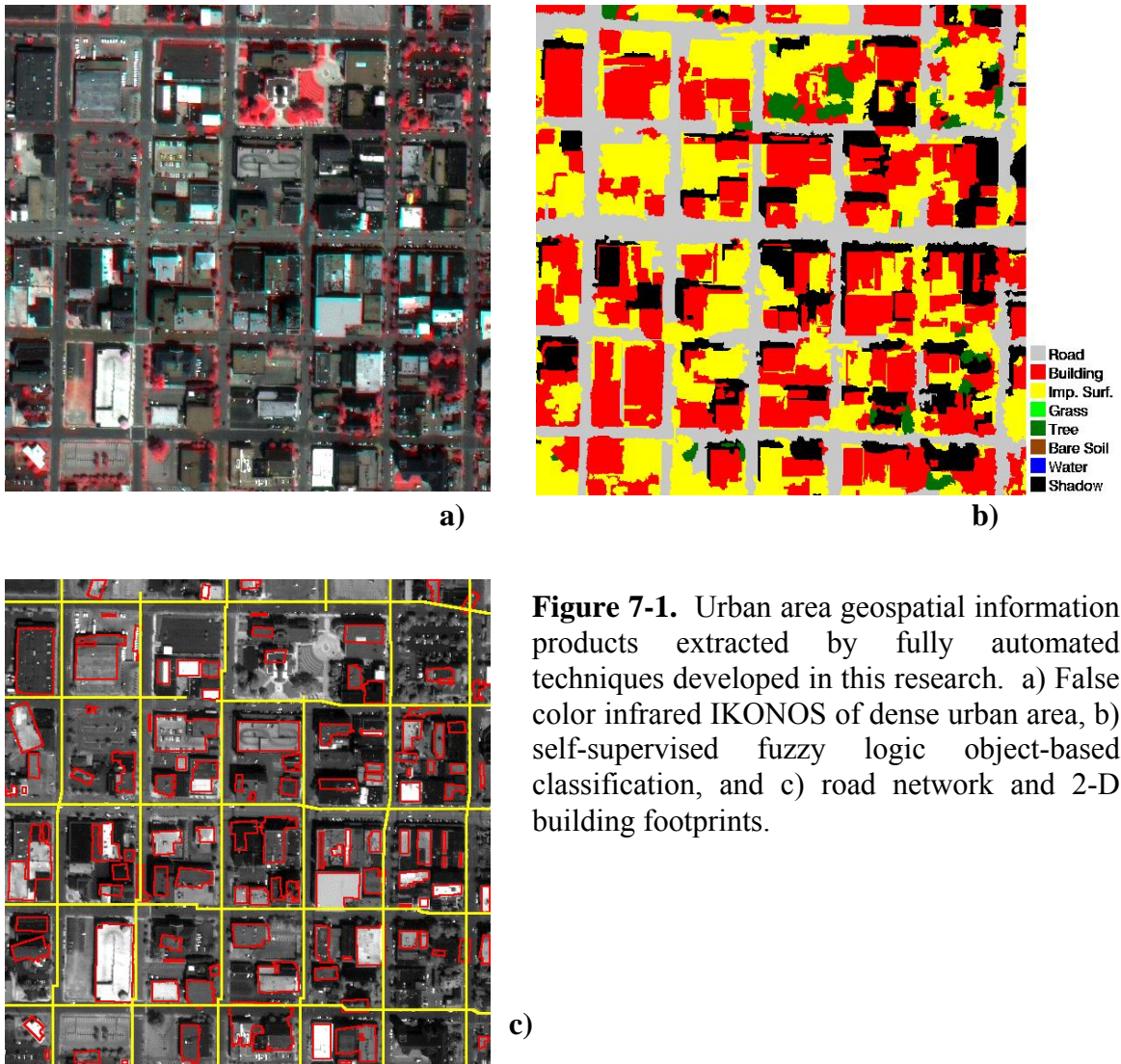
## **Chapter 7**

### **Summary and Future Work**

#### **7.1 Research Summary**

The goal of this doctoral research was to develop advanced automated and semi-automated image analysis and classification techniques for the extraction of urban area geospatial information products from commercially available high-resolution satellite imagery. Specifically, the problems of both semi and fully automated urban land cover classification and automated extraction of urban infrastructure features, e.g. road networks and 2-D building footprints, were examined. The urban area geospatial information products extracted utilizing the fully automated techniques developed in this doctoral research are shown in Fig. 7-1. The development of automated and semi-automated techniques for generation of urban geospatial information products is of high importance not only because of the many applications for which they can be used, but also because the large volume of data collected by satellite sensors exceeds the human capacity of trained image specialists to analyze. Many applications, especially military and intelligence related activities, require near real time exploitation of image data. Both the quantity of data and the speed with which that data must be analyzed illustrate the need for automated geospatial information extraction. The spatial resolution of commercially available satellite imagery did not increase to the level of detail required for most urban area applications until recent years (2000). As such, relatively little work

has been reported in the literature on the exploitation of this type of data for urban area applications.



**Figure 7-1.** Urban area geospatial information products extracted by fully automated techniques developed in this research. a) False color infrared IKONOS of dense urban area, b) self-supervised fuzzy logic object-based classification, and c) road network and 2-D building footprints.

In Chapter 1, a variety of background material was presented on geospatial information products, high-resolution satellite imagery, and image processing techniques and methods. The utility of these products for urban area applications at both the local and federal government levels, as well as with the intelligence and military communities, was discussed. Several earth image collection methods were discussed, including both

aerial photography and satellite based digital sensors. In particular, high-resolution imagery from commercial satellites such as IKONOS and QuickBird were shown to be well suited for use in many urban area applications. Following the discussion of the image sensor systems, a number of relevant image analysis techniques were reviewed. Both supervised and unsupervised classification techniques were reviewed, and several techniques, such as maximum likelihood, were discussed in detail. This was followed by an introduction to fuzzy set theory, fuzzy logic, and their application to remote sensing image analysis. Finally, a summary of related work in the areas of classification techniques, segmentation techniques, analysis of urban areas, automated road extraction, and automated building extraction was presented.

The problem of per-pixel supervised classification of high-resolution satellite multispectral (MS) imagery of urban areas was addressed in Chapter 2. The results presented in this chapter demonstrated the usefulness of high-resolution satellite imagery for urban land cover mapping and some of the shortcomings of conventional classification techniques such as maximum likelihood. The urban land cover classes used were *Road*, *Building*, *Grass*, *Tree*, *Bare Soil*, *Water*, and *Shadow*. It was found that maximum likelihood classification of high-resolution MS imagery over urban areas produced significant misclassification errors between spectrally similar classes. Different spatial measures such as texture and contextual methods were investigated and found to increase the discrimination between certain spectrally similar classes. A hierarchical fuzzy classification method was developed that utilized both the spectral and spatial information to classify the data. The fuzzy classification approach is based on a framework that allows for different groups of classes to be classified using the features

best suited for discrimination between those classes. This alleviates the problem of features simultaneously decreasing the confusion between one set of classes and increasing it for another set. The classification accuracies of the fuzzy classifier were approximately 10% greater than the maximum likelihood classification results. Accordingly, there were significant decreases in the number of misclassifications between spectrally similar classes.

A combined fuzzy pixel-based and object-based approach for urban land cover classification of high-resolution MS image data was presented in Chapter 3. In addition to the urban land cover classes identified by the fuzzy pixel-based classifier, a non-road, non-building *Impervious Surface* class was added to further refine the land cover classification in dense urban areas. To distinguish between these urban land cover classes, an object-based classification approach was introduced to examine features such as object shape and neighborhood and then classify the image objects using a fuzzy logic rule base. To facilitate object classification, the imagery was first segmented with a region merging segmentation technique. Object level features were extracted from the image objects and used by the object-based classifier along with the fuzzy pixel-based classification. These features are the class labels of each segment's constituent pixels, shape information from the image objects, neighborhood analysis, and spectral statistics of the object. The image objects were then classified with a fuzzy logic rule base. This classification approach produced an urban land cover map with a 76% accurate *Building* class and an 81% accurate *Impervious Surface* class.

The problem of automated extraction of the road network from high-resolution imagery over urban areas was examined in Chapter 4. Two techniques for extracting



roads from land cover classification maps were presented: 1) skeletonization and 2) iterative line segment identification and growth algorithm. Skeletonization was found to produce large numbers of false positive road identifications. The road network extracted by the iterative algorithm was accurate (87-93% correctness), however it was only able to produce a partial extraction resulting in completeness values of 70%. Following the techniques for road extraction from land cover classifications, a fully automated road network extraction algorithm for high-resolution multispectral data over urban areas was presented. Roads were extracted using 2-D spatial feature information after masking out vegetation pixels using the NDVI statistic. An iterative road segment identification and growth algorithm was then introduced. This algorithm iteratively added road segments to the road network, beginning with the longest length line segments in the imagery and progressing to segments of shorter length. The endpoints of the road segments were grown so curvilinear roads could be extracted. A buffer around the road network was utilized to keep spurious line segments from being added to the existing road network. The extraction algorithm was tested on several sites from an IKONOS high-resolution satellite image and digital aerial orthophotographs. The algorithm was shown to produce road networks that were 79–87% complete and 70–87% correct.

A fully automated algorithm for 2-D building footprint extraction from high-resolution MS satellite imagery was presented in Chapter 5. The algorithm was based on a multi-detector fusion strategy where buildings and their shadows were extracted from the Differential Morphological Profile (DMP) of panchromatic imagery and a segmentation of the pan-sharpened multispectral imagery. The DMP is a multi-scale image analysis technique that provides information about both the size and contrast of

multi-scale structures in the image. Buildings with a bright spectral response were extracted from the opening portion of the differential profile, while shadows were extracted from the closing portion of the differential profile. The extracted shadows define a search area where the presence of a building is likely, which is then utilized to identify building objects from the segmented PS-MS image. The algorithm was tested on a subset of the Columbia IKONOS imagery containing dense urban land cover and was found to produce an accurate (correctness values of 87% and 77% for the object and pixel assessments, respectively) yet somewhat incomplete (completeness values of 71% and 50% for the object and pixel assessments, respectively) building footprint extraction.

A fully automated self-supervised approach for urban land cover classification of high-resolution MS satellite imagery was presented in Chapter 6. The classifier was based on the supervised classification approaches developed previously in this research (Chapters 2 and 3). However, the classifier does not require manually identified training data. Instead, the training data is automatically generated internally using feature extraction techniques that identify high confidence instances of the urban land cover classes. The automated road network and 2-D building footprint extraction algorithms presented in Chapters 4 and 5 were modified to output fuzzy confidence values for each extracted feature. Additional spatially and spectrally based feature extraction algorithms were developed to identify training data for the other urban land cover classes. After feature extraction and selection of high accuracy training data, the extracted features are subdivided into spectrally coherent sub-classes by unsupervised spectral clustering. This internally generated training data was then used to train a maximum likelihood classifier, followed by the hierarchical fuzzy pixel-based classifier, and finally the object-based

classifier. Test results indicate that the fully automated self-supervised classification approach is able to produce urban land cover maps with overall accuracies that are only 2-3% less than that of the manually trained supervised classifiers.

The classification techniques discussed in Chapter 2 were presented at the *2002 International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2002] and were published in the *IEEE Transactions on Geoscience and Remote Sensing - Special Issue on Urban Satellite Remote Sensing* [Shackelford and Davis, 2003c]. The object-based classification approach presented in Chapter 3 was published in the *IEEE Transactions on Geoscience and Remote Sensing* [Shackelford and Davis, 2003d]. The material presented in Chapter 4 on automated road network extraction was presented at the *2003 GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas* [Shackelford and Davis, 2003a] and the *2003 International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2003b]. The material presented in Chapter 5 on automated 2-D building footprint extraction was presented at the *2004 International Geoscience and Remote Sensing Symposium* [Shackelford and Davis, 2004]. The material presented in Chapter 6 on fully automated self-supervised urban land cover classification will be submitted to the *IEEE Transactions on Geoscience and Remote Sensing* for publication.

## **7.2 Future Work**

A variety of techniques for automated extraction of urban area geospatial information products from high-resolution satellite imagery have been developed in this doctoral research. There are a number of areas in which future research, building on the

results presented here, can be explored. For example, many of the developed techniques make use of fuzzy membership functions to model the uncertainty and ambiguity associated with the shape and spectral responses of different urban land cover classes and features. In a number of cases, the parameters governing the shape of these membership functions have been manually set. Similarly, the operation of the road network and building footprint extraction algorithms are controlled by a number of manually set parameters. While these parameters are intuitively chosen based on characteristics of the imagery and the type of feature they are to represent, it would be beneficial to develop techniques for learning them directly from the data. One possible approach would be to utilize neural network type techniques, as was done in Chapter 2 to learn the shape of the membership functions for the *Road* and *Building* classes from the 2-D spatial contextual features.

The use of more advanced data fusion and fuzzy membership combination techniques, such as the fuzzy integral, should be explored. Currently, fuzzy membership values are combined using the standard fuzzy union and intersection operators, min and max. The fuzzy integral is an important technique for combining information from multiple data sources. Information is fused by combining objective evidence from each information source with subjective evaluation of the importance or worth of the sources [Bezdek *et al.*, 1999]. The subjective assessment of “worth” can be based on known confidences in a given source/method, human intuition, or on empirical data. By utilizing techniques such as the fuzzy integral, multiple features can be combined in a more intuitive manner. For example, if during the extraction of a building each of the multiple sources of evidence indicates only a medium level of confidence in the presence of a

building, the fuzzy integral can be used to combine the evidence in a meaningful way increasing the confidence in the presence of a building.

The use of the DMP for the purpose of image segmentation, as proposed by Pesaresi and Benediktsson [2001], should be explored. As with image objects generated by MS segmentation, spectral, spatial, and contextual features can be calculated for each DMP image object. Multiple sets of image objects can be generated using different segmentation techniques, and features extracted from the image objects can be combined using approaches such as the fuzzy integral. The fusion of multiple sets of independent image objects could help mitigate problems of over and under segmentation. The object-based features can then be used to supplement the pixel-based features as additional data bands in a supervised classification system, possibly increasing the ability of the classifier to discriminate between spectrally similar classes. Another source of information that could be exploited by a classifier are the fuzzy confidence values of extracted urban land cover features. In addition to object-based features, these fuzzy confidence values could be added as additional bands of data, to help decrease the number of misclassifications.

The pixel-based hierarchical fuzzy classification scheme discussed in Chapter 1 employs maximum likelihood classification to divide the input data into sets of spectrally similar classes for the fuzzy classifier to operate on. Because maximum likelihood classification assumes normally distributed data, features such as the 2-D spatial contextual features, which are not normally distributed, cannot be utilized with this type of classifier. Other classification algorithms that require no assumptions about the distribution of the data, such as support vector machines, should be investigated for use

with features of this type. Support vector machines perform a nonlinear mapping of the input data to high-dimensional feature space, where optimal decision hyperplanes are constructed between a subset of the training data from the different classes. Support vector machines are well suited to situations where no assumption about the distribution of the input data is made, or the input data is high dimensional. For a detailed description of support vector machines, see Haykin [1999].

Finally, the techniques developed in this research should be adapted for use on the next generation satellite imagery data. Proposed systems, such as DigitalGlobe's WorldView, will have sub-meter spatial resolution and additional MS bands, extending into the low end of the infrared spectrum. These characteristics will allow for finer detail mapping of urban structures and possibly the identification of additional land cover classes due to the increase in spectral resolution. The additional spectral information will likely have a significant impact on the ability of supervised and self-supervised classification systems to accurately classify urban areas.

## Reference List

- Agouris, P., P. Doucette and A. Stefanidis, 2001, "Spatiospectral cluster analysis of elongated regions in aerial imagery," *Proc.of IEEE International Conference on Image Processing*, Vol. 1, pp. 177-180, October, 2001.
- Baatz, M., and A. Schape, 2000, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informationsverarbeitung. XII. Beitrage zum AGIT-Symposium Salzburg 2000*, T. Strobl, T. Blaschke, and G. Griesebner, Eds., Karlsruhe: H. Wichmann Verlag, pp. 12-23.
- Bardossy, A., and L. Samaniego, 2002, "Fuzzy rule-based classification of remotely sensed imagery," *IEEE Trans. Geosci. Remote Sensing*, Vol. 40, No. 2, pp. 362-374.
- Barnes, S., 2003, "DigitalGlobe lands \$500 million NextView contract," *Geospatial Solutions*, November 2003, pp. 14-15.
- Baumgartner, A., C. Steger, H. Mayer, W. Eckstein and H. Ebner, 1999, "Automatic road extraction based on multi-scale, grouping, and context," *Photogrammetric Engineering & Remote Sensing*, Vol. 65, No. 7, pp. 777-785.
- Benediktsson, J. A., M. Pesaresi, and K. Arnason, 2003, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sensing*, Vol. 41, No. 9, pp. 1940-1949.
- Bezdez, J. C., J. Keller, R. Krisnapuram and N. R. Pal, 1999, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic Publishers: Boston.
- Campbell, J. B., 1996, *Introduction to Remote Sensing*, 2<sup>nd</sup> ed., Guilford Press: New York.
- Congalton, R. G., R. G. Oderwald, and R. A. Mead, 1983, "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques," *Photogrammetric Engineering & Remote Sensing*, Vol. 49, No. 12, pp. 1671-1678.
- Congalton, R. G., 1991, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, Vol. 37, pp. 35-46.
- Couloigner, I., and T. Ranchin, 2000, "Mapping of urban areas: a multiresolution modeling approach for semi-automatic extraction of streets," *Photogrammetric Engineering & Remote Sensing*, Vol. 66, No. 7, pp. 867-874.

- Croitoru, A. and Y. Doytsher, 2003, "Monocular right-angle building hypothesis generation in regularized urban areas by pose clustering," *Photogrammetric Engineering & Remote Sensing*, Vol. 69, No. 2, pp. 151-169.
- Davis, C. H., and X. Wang, 2003, "Planimetric accuracy of Ikonos 1-m panchromatic orthoimage products and their utility for local government GIS basemap applications," *International Journal of Remote Sensing*, Vol. 24, No. 22, pp. 4267-4288.
- Dell'Acqua, F., and P. Gamba, 2001, "Detection of urban structures in SAR images by robust fuzzy clustering algorithms: the example of street tracking," *IEEE Trans. Geosci. Remote Sensing*, Vol. 39, No. 10, pp. 2287-2297, 2001.
- DigitalGlobe, 2003, "QuickBird Imagery Products: Product Guide – Revision 3.8," – <http://www.digitalglobe.com/downloads/QuickBird%20Imagery%20Products%20-%20Product%20Guide.pdf>
- Doucette, P., P. Agouris, A. Stefanidis and M. Musavi, 2001, "Self-organized clustering for road extraction in classified imagery," *ISPRS Journal of Photogrammetry & Remote Sensing*, Vol. 55, pp. 347-358.
- Evans, C., R. Jones, I. Svalbe, and M Berman, 2002, "Segmenting Multispectral Landsat TM images into field units," *IEEE Trans. Geosci. Remote Sensing*, Vol. 40, No. 5, pp. 1054-1064.
- Ferro, C. J. S., and T.A. Warner, 2002, "Scale and texture in digital image classification," *Photogrammetric Engineering & Remote Sensing*, Vol. 68, No. 1, pp. 51-63.
- Gonzalez, R. C. and R. E. Woods, 2002, *Digital Image Processing*, 2<sup>nd</sup> ed., Prentice Hall: New Jersey.
- Haralik, R. M., K. Shanmugam, and D. Its'hak, 1973, "Textural features for image classification," *IEEE Trans. Syst. Man Cyber.*, Vol. SMC-3, pp. 610-621.
- Haverkamp, D., 2002, "Complementary methods for extracting road centerlines from IKONOS imagery," *Proceedings of SPIE 9<sup>th</sup> International Symposium on Remote Sensing*, pp. 501-511, Crete, Greece, 22-27 September 2002.
- Heipke, C., H. Mayer, C. Wiedemann, and O. Jamet, 1997, "Evaluation of automatic road extraction," *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, No. 3, pp. 47-56.
- Hinz, S., and A. Baumgartner, 2003, "Multi-view fusion of road objects supported by self-diagnosis," *Proceeding of 2<sup>nd</sup> GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp. 137-141, Berlin, 22-23 May, 2003.



- Hudson, W. D. and C. W. Ramm, 1987, "Correct formulation of the kappa coefficient of agreement," *Photogramm. Eng. Remote Sens.*, Vol. 53, No. 4, pp. 421-422, Apr. 1987.
- Haykin, S., 1999, *Neural Networks: A Comprehensive Foundation*. Prentice Hall: New Jersey.
- Irvin, R. B., and D. M. McKeown, 1989, "Methods for exploiting the relationship between buildings and their shadows in aerial imagery," *IEEE Trans. Syst. Man Cyb.*, Vol. 19, No. 6 pp. 1564-1575.
- Jang, B. K. and R. T. Chin, 1990, "Analysis of thinning algorithms using mathematical morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 12, No. 6, pp. 541-551.
- Jaynes, C. O., F. Stolle and R. T. Collins, 1994, "Task driven perceptual organization for extraction of rooftop polygons," *IEEE Workshop on Applications of Computer Vision*, pp. 152-159, December 1994.
- Jenson, J. R., 1996, *Introductory Digital Image Processing*, 2<sup>nd</sup> ed., Prentice Hall: New Jersey.
- Jenson, J. R. and D. C. Cowen, 1999, "Remote sensing of urban/suburban infrastructure and socio-economic attributes," *Photogrammetric Engineering & Remote Sensing*, Vol. 65, No. 5, pp. 611-612.
- Klir, G. J., and B. Yuan, 1995, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall: New Jersey.
- Kressler, F. P., T. B. Bauer, and K. T. Steinnocher, 2001, "Object-oriented per-parcel land use classification of very high resolution images," *Proceeding of IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Rome, Italy*, pp. 164-167, Nov. 2001.
- Lee, D. S., J. Shan and J. S. Bethel, 2003, "Class-guided building extraction from IKONOS imagery," *Photogrammetric Engineering & Remote Sensing*, Vol. 69, No. 2, pp. 143-150.
- Light, D.L, 1993, "The National Aerial Photography Program as a Geographic Information System Resource," *Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 1, pp. 61-65.
- Longley, P. A., M. F. Goodchild, D. J. Maguire and D. W. Rhind, 2001, *Geographic Information Systems and Science*, Wiley: New York.
- Melgani, F., B. A. R. AL Hashemy, and S. M. R. Taha, 2000, "An explicit fuzzy supervised classification method for multispectral remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, Vol. 38, No. 1, pp. 287-295.

Mayer, H., I. Laptev, A. Baumgartner and C. Steger, 1997, "Automatic road extraction based on multi-scale modeling, context, and snakes," *International Archives of Photogrammetry & Remote Sensing*, Vol. 32, 3-2W3, pp. 106-113.

National Imagery and Mapping Agency, 2003, "Geospatial intelligence: capstone concept," - <http://www.fas.org/irp/agency/nima/capstone.pdf>.

Pesaresi, M., 1999, "Textural classification of very high-resolution satellite imagery: empirical estimation of the interaction between window size and detection accuracy in urban environment," *Proceedings of IEEE International Conference on Image Processing*, Vol. 1, pp. 114-118, Kobe, Japan, October, 1999.

Pesaresi, M., and J.A. Benediktsson, 2001, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sensing*, Vol. 39, No. 2, pp. 309-320.

Ranchin, T. and L. Wald, 2000, "Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation," *Photogrammetric Engineering & Remote Sensing*, Vol. 66, No. 1, pp. 49-61.

Sarkar, A., M.K. Biswas, B. Kartikeyan, V. Kumar, K.L. Majumder, and D.K. Pal, 2002, "A MRF model – based segmentation approach to classification for multispectral imagery," *IEEE Trans. Geosci. Remote Sensing*, Vol. 40, No. 5, pp. 1102-1113.

Segl, K. and H. Kaufmann, 2001, "Detection of small objects from high-resolution panchromatic satellite imagery based on supervised image segmentation," *IEEE Trans. Geosci. Remote Sensing*, Vol. 39, No. 9, pp. 2080-2083.

Shackelford, A. K. and C.H. Davis, 2002, "A fuzzy classification approach for high-resolution multispectral data over urban areas," *Proceedings of International Geoscience and Remote Sensing Symposium*, Vol. 3, pp. 1621-1623, Toronto, Canada, 24-28 June, 2002.

Shackelford, A. K. and C. H. Davis, 2003a, "Urban road network extraction from high-resolution multispectral data," *Proceeding of 2<sup>nd</sup> GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp. 142-146, Berlin, 22-23 May 2003.

Shackelford, A. K. and C. H. Davis, 2003b, "Fully automated road network extraction from high-resolution satellite multispectral imagery," *Proceedings of International Geoscience and Remote Sensing Symposium*, Vol. 1, pp. 461-463, Toulouse, France, 21-25 July, 2003.

Shackelford, A. K. and C. H. Davis, 2003c, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sensing*, Vol. 41, No. 9, pp. 1920-1932.

- Shackelford, A. K. and C. H. Davis, 2003d, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sensing*, Vol. 41, No. 10, pp. 2354-2363.
- Shackelford, A. K., C. H. Davis, and X. Wang, 2004, "Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery," *Proceedings of International Geoscience and Remote Sensing Symposium*, Vol. 3, pp. 1996-1999, Anchorage, Alaska, 20-24 September, 2004.
- Shi, W., and C. Zhu, 2002, "The line segment match method for extracting road network from high-resolution satellite images," *IEEE Trans. Geosci. Remote Sensing*, Vol. 40, No. 2, pp. 511-514.
- Soille, P., and M. Pesaresi, 2002, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sensing*, Vol. 40, No. 9, pp. 2042-2055.
- Space Imaging, 2003, "IKONOS Product Guide – Version 1.3," - [http://www.spaceimaging.com/whitepapers\\_pdfs/IKONOS\\_Product\\_Guide.pdf](http://www.spaceimaging.com/whitepapers_pdfs/IKONOS_Product_Guide.pdf).
- Steger, C., 1998, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 20, No. 2, pp. 113-125.
- Tatem, A. J., H. G. Lewis, P. M. Atkinson, and M. S. Nixon, 2001, "Super-resolution mapping of urban scenes from IKONOS imagery using a Hopfield neural network," *Proceedings of International Geoscience and Remote Sensing Symposium*, Vol. 7, pp. 3203-3205.
- Tilton, J. C., 1998, "Image segmentation by region growing and spectral clustering with a natural convergence criterion," *Proceedings of International Geoscience and Remote Sensing Symposium*, Vol. 4, pp. 1766-1768, Seattle, WA, July 6-10, 1998.
- Theodoridis, S. and K. Koutroumbas, 1998, *Pattern Recognition*, Academic Press: California.
- United States Geological Survey, 1996, "Standards for Digital Orthophotos, Part 1: General," U.S. Department of the Interior, Washington, DC.
- United States Geological Survey, 1999, "Landsat 7 Level-0 and Level-1 Datasets Document," U.S. Department of the Interior, Washington, DC - [http://eosims.cr.usgs.gov:5725/DATASET\\_DOCS/landsat7\\_dataset.html](http://eosims.cr.usgs.gov:5725/DATASET_DOCS/landsat7_dataset.html).
- United States Geological Survey, 2001, "The National Map," USGS Fact Sheet 101-01, U.S. Department of the Interior, Washington, DC.

United States Geological Survey, 2003, "Landsat: A Global Land-Observing Program," USGS Fact Sheet 023-03, U.S. Department of the Interior, Washington, DC.

Van Teeffelen, P., S. de Jong, and L. van der Berg, 2001, "Urban monitoring: new possibilities of combining high spatial resolution IKONOS images with contextual image analysis techniques," *Proceeding of IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, Rome, Italy, pp. 265-269, Nov. 2001.

Vrabel, J., 1996, "Multispectral imagery band sharpening study," *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 9, pp. 1075-1083.

Wang, F., 1990, "Fuzzy supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sensing*, Vol. 28, No. 2, pp. 194-201.

## VITA

Aaron Kenneth Shackelford was born in Kansas City, Missouri on January 1, 1977. He received the B.S. (with honors), M.S., and Ph.D. degrees in Electrical Engineering from the University of Missouri-Columbia in 1999, 2001 and 2004, respectively. Since 2000, he has been a Graduate Research Assistant in the Department of Electrical and Computer Engineering at the University of Missouri-Columbia. He was a Research Scholar in the Department of Electronic Engineering, City University of Hong Kong, during the summer of 2000, where he conducted research on microstrip patch antennas. He was awarded the NASA Graduate Student Researcher Program Fellowship in 2001 to pursue his Ph.D. studies. His doctoral research focused on the development of techniques for automated extraction of urban area geospatial information products from high-resolution satellite imagery using advanced image processing techniques and information fusion strategies. His work was featured on the cover of the September 2003 issue of the *IEEE Transactions on Geoscience and Remote Sensing Special Issue: Urban Remote Sensing by Satellite*. In 2000, he married the former Lisa Wilson of Kansas City, Missouri.