Aerospace Engineering Sciences Graduate Theses & Dissertations

Aerospace Engineering Sciences

Spring 1-1-2012

Urban Image Classification Using Multi-Angle Very-High Resolution Satellite Data

Nathan W. Longbotham
University of Colorado at Boulder, nathan.longbotham@colorado.edu

Urban Image Classification Using Multi-Angle
Very-High Resolution Satellite Data

by

Nathan. W. Longbotham

B.S., Abilene Christian University, 2001

M.S., University of New Mexico, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Aerospace Engineering Sciences

2012

This thesis entitled:
Urban Image Classification Using Multi-Angle
Very-High Resolution Satellite Data
written by Nathan. W. Longbotham
has been approved for the Department of Aerospace Engineering Sciences

William Emery

James Maslanik

Robert Leben

Cora Randall

Kumar Navulur

Date

The final copy of this thesis has been examined by the signatories, and we find that both the
content and the form meet acceptable presentation standards of scholarly work in the above
mentioned discipline.

Longbotham, Nathan. W. (Ph.D., Aerospace Engineering Sciences)

Urban Image Classification Using Multi-Angle

   Very-High Resolution Satellite Data

Thesis directed by Prof. William Emery

The ability to automatically generate large-area land-use/land-cover (LU/LC) classification maps from very-high spatial resolution (VHR) satellite data is dependent on two capabilities: (1) the ability to create a data model able to accurately classify satellite data into the appropriate surface types and (2) the ability to apply this model to the multiple images necessary to create a large-area VHR mosaic. This research describes methods for improving these capabilities by leveraging the unique characteristics of VHR in-track and composite multi-angle data.

It is shown that new features can be extracted from both in-track and composite multi-angle data in order to improve classification performance. These features encode information extracted from the spatial and spectral variations of the multi-angle data, such as spectral fluctuation with view-angle and pixel height. This additional knowledge provides the capability to both improve image classification performance (29% in the demonstrated experiments) and include urban LU/LC classes, such as bridges, high-volume highways, and parking lots, that are normally difficult to identify in multispectral urban data.

Additionally, methods that apply a multispectral classification model across multiple images (model portability) are also explored using the simplifying test cases of in-track and composite multi-angle data. The in-track results show that the portability of a multispectral model can be improved from no portability (losing all classification capability when applying the model across the multi-angle images) to a 10% reduction in kappa coefficient across the sequence of in-track images when physically based image normalization techniques are appropriately applied. The additional noise of seasonality limits the portability performance in the composite multi-angle sequence to an approximate reduction in kappa coefficient of 20% in the best cases.

Dedication

To my family...

To my wife, Jennifer, thank you for the push to start this process and for the continued support through the stress and long nights of its completion. To my parents, Jerry and Lee Ann Longbotham, and siblings, Tammy Barone, Lori Hooten, and Andrew Longbotham - your unwavering support and confidence has been priceless and a source of continued strength.

Acknowledgements

Contents

Appendix

Tables

Table

Figures

Figure

Chapter 1

Introduction

Remote sensing plays a unique role in the areas of public policy and scientific exploration by providing large scale information that would otherwise be unattainable. The tools developed by the remote sensing community enable Earth observation and monitoring for many areas of importance, including human settlement dynamics, environmental changes, and economics. This additional information can fundamentally change the dynamics of a policy decision or scientific investigation by extending the discussion from one based on data limited in both temporal and spatial extent, to a more informed analysis based on extended observations.

Information extraction from satellite imagery provides a unique opportunity to create this type of extended information, as satellites operate independently from ground-based infrastructure and can collect data over in-accessible areas with frequent re-visits to provide dense sampling (spatial, temporal, spectral, and angular) of the target area. The information extracted from these data would otherwise be unavailable and, therefore, is of unique value. It is precisely this capability - the ability to densely sample a target area in a variety of modalities and extract actionable information - that drives the academic research in this area.

Currently, the majority of satellite imagery analysis is carried out manually by expert operators using visual interpretation to extract the information of interest within the acquired data. However, as more satellites come into operation with increasing spatial, spectral, and temporal resolutions, the volume of data is increasing dramatically, and, therefore, the human analysis process becomes increasingly untenable.

In order to address this limitation, automated processes are increasingly employed to handle the transformation of this large volume of data into analyzed information. While the desired end product can take many forms, the required information retrieval generally falls into one of three categories:

(1) Land-Use/Land-Cover (LU/LC) Classification: Creation of maps that describe the spatial structure and distribution of land-cover types (asphalt, trees, grass, etc.) and/or land-use types (parking lots, highways, residential homes, etc.) within a target area.

(2) Change Detection: Identification of places within a target area that have changed between two or more observations.

(3) Object Detection: Extraction and enumeration of individual objects (vehicles, trees, airplane, etc.) from a target area.

The research presented in this dissertation focuses on the first of the information extraction techniques, specifically, on large-area urban LU/LC classification of very-high spatial resolution (VHR) satellite data.

The development of land-use/land-cover classification has been identified as a research area of significant importance in the coming decades. It is one of six "grand challenges" of the National Ecological Observatory Network (NEON), a distributed observatory designed to address areas of need in environmental science with the goal of better understanding climate dynamics on a continental scale [4]. This focus is due to its capability to significantly improve both the temporal and spatial information available to scientific studies through automated means of identifying surface cover. For example, in the monitoring of invasive species or seasonal ecological cycles, automated methods to identify and quantify the extents and temporal patterns of various ground cover types can be invaluable.

This type of monitoring can also be extended to urban areas, as is the focus of this research. The most often cited use of urban classification is urban planning and monitoring. While this is indeed an interesting application with benefit to activities such as development monitoring and

impervious surface mapping, there are additional applications that drive investigations in multiple scientific and engineering fields. These include urban forest mapping, urban heat islands measurements, and mapping of economic indicators.

Urban areas are of particular interest in this field because the spatial and spectral complexity of a typical urban area make automated methods of information extraction difficult. This is due, in part, to the limited spatial resolution of widely available moderate resolution (100 m to 1 km) satellite imagery. With the relatively recent availability of VHR satellite data (approximately one meter spatial resolution), satellite images are more readily available that clearly depict the meter scale features that are integral to urban areas.

With the additional spatial resolution afforded by VHR satellite data, it is possible to develop automated techniques that more accurately handle the spatial and spectral complexities of urban areas. The techniques required for the creation of large-area urban LU/LC classification maps broadly encompass two capabilities:

(1) LU/LC classification: The ability to create a data model able to accurately classify satellite data into the appropriate surface types.

(2) LU/LC model portability: The ability to use these models on the multiple images necessary to create a large-area VHR image mosaic.

This research explores methods to address and improve these capabilities by leveraging the unique characteristics of two types of VHR multi-angle data: in-track and composite. In-track data is composed of multiple images collected sequentially during a single satellite over-flight. This type of multi-angle data encompasses both the angular dependences of surface reflectance as well as short term (seconds) surface changes. Composite multi-angle data is a sequence of multi-temporal images with each image collected at a different view-angle. This type of multi-angle data set encompasses both the angular dependences of surface reflectance as well as long-term (days and months) surface changes. In both cases, the data is collected by the WorldView-2 satellite with 50 cm panchromatic

resolution and 2 m resolution in eight multispectral optical bands. These data sets are discussed further in Chapter 3.

1.1 Land-Use/Land-Cover Classification

The first requirement for the development of large-area LU/LC classification maps, described above, is a data model capable of accurately describing the surface materials of interest from the information available. In this dissertation, this requirement is explored by using the information sources present within multi-angle VHR optical satellite imagery to significantly improve LU/LC classification models.

Multispectral VHR imagery is uniquely suited to LU/LC classification because the imagery provides the necessary detail to depict typical urban and suburban features. This data also typically includes multispectral information in the optical wavelengths as well as an additional panchromatic band at higher spatial resolution observing an integrated solar radiance over the optical wavelengths. This information forms the foundation of typical LU/LC classification in that it provides the spectral information necessary to differentiate basic classes, such as vegetation, soil, water, and shadow.

However, the spectral similarity of many land-cover types limit the utility of raw spectral information as for urban LU/LC classification. Surfaces that have little, if any, differences in spectral signature between them, such as parking lots and highways or sidewalks and commercial buildings, can not be distinguished using spectral information alone. In order to increase the performance of LU/LC classification, new capabilities must be brought to the problem to differentiate sources with spectral similarity. This can take several forms:

(1) Improve Model Algorithms

(2) Leverage more data

(3) Instrument more information

(4) Compute Contextual Features

(5) Fuse existing data

Algorithms: Many machine learning algorithms have been developed with the goal of improving the utilization of classification information. Generally, the approach involves creating classes that can more accurately conform and segment the nonlinearity of an arbitrary data space [72,95]. However, it has recently been argued that improvements in machine learning algorithms have begun to wane [109]. Conversely, some techniques seek to develop algorithms that attempt to linearize a non-linear data space [19] such that it can more easily be classified by simple linear statistical data model.

High-Performance Computing: Increasingly accessible large computational systems improve the ability to analyze an ever expanding amount of remotely-sensed data [82]. These systems generally improve the performance of remote sensing analyses in two ways: by enabling computationally complex feature extraction methods and by increasing the amount of data that can be included into an analysis system. The ability to perform feature extraction on large data sets, and then use the extracted information, is a key capability for remote sensing technologies. These applications typically consist of implementing well known, but computationally intense, techniques for the creation of unique data features. [58] However, the sheer size of a large-area remote sensing data set requires significant computational resources. This is increasingly true when the additional complexity of derived features is added. By leveraging the power of large computational systems, it is possible to extend remote sensing analyses into larger data domains (larger area, finer spatial resolution, more spectral information, extracted data features etc.) that more accurately represent the data space in question [85].

Instrumentation: It is also possible to increase the utility and accuracy of LU/LC classification by using new instrumentation. By creating sensors that can provide more information about a given problem, it is possible to outperform more traditional optical sensors by including the new observations of the target of interest. This additional instrumented data can take many forms, including increased spatial resolution, extended spectral observations, and waveform measurements.

This type of information is collected by sensors such as waveform LIDAR (LIght Detection And Ranging), which can described volumetric data such as forest bio-mass density [110], synthetic aperture radar (SAR), which records Doppler and backscatter data in the radar region of the electromagnetic spectrum [95], and first-return LIDAR, for measuring surface height [25]. In this way, improvements are made in remote sensing analyses through a more extensive description of the full physical space occupied by a target of interest.

Compute Contextual Features: The creation of image features is essential to the analysis of typical RGB (red, green, and blue) pictures. The human visual system (HVS) consists of three RGB spectral bands that people use for the identification of objects in an image. Although the basis of the HVS is the raw spectral information, humans have no problem identifying (at the appropriate spatial resolution) whether a red pixel is a car or a roof top. That is because humans instinctively query surrounding information when considering an image object.

The development of this type of information, created from a single observation, can take many forms. These techniques include object based statistics, texture, and morphology, and have provided significant improvement for LU/LC classification accuracy [79, 104]. While most feature creation techniques involve the derivation of contextual information in ways similar to that in which humans interpret photographs, more recent methods develop contextual information in the data space (as opposed to the spatial structure of the image) through the use of techniques such as statistical modeling [92, 112] and graph theory [76].

Data Fusion: The combination of complementary, and spatially overlapping, data from separate observations to add information to a remote sensing analysis problem is referred to as data fusion [113]. Classically this information comes from different platforms - such as fusing optical and SAR observations [67]. However, rarely does simply combining multiple data sets improve the classification problem of interest. More often, it is necessary to overcome and leverage the temporal, spatial, and spectral differences between the data sets or to create a mechanism by which new information features can be extracted from the fused data. While powerful in many applica-

tions, data fusion comes at the cost of operating multiple instruments and implementing methods of integrating instrumentation.

The classification research presented in this dissertation focuses on developing additional information from the fusion of multi-angle images in order to improve LU/LC classification. Specifically, classification is improved by using multiple WorldView-2 observations to extract height, multi-angle spectral, and multi-temporal information from multiple images collected by a single optical platform. The methods, difficulties, and results of fusing multiple observations from two types of multi-angle data (in-track and composite - see Chapter 3) for the purposes of improved LU/LC classification, will be shown in Chapter 6.

1.2   Model Portability

The spatial detail available from VHR satellite imagery is particularly useful for creating LU/LC classification maps because the resolution allows for identification of individual objects, such as trees, roadways, and homes, found in urban areas. However, the swath width (the observation width of the satellite instrument) of VHR satellite systems is relatively limited when compared to moderate and low spatial resolution satellite systems. Therefore, to create large-area LU/LC classification maps from VHR data, it is necessary to use data from multiple images.

Therefore, in order to develop large-area LU/LC classification maps, a data model must be applicable to the multiple images required for a large-area VHR mosaic. In this dissertation, foundational research is reported on data normalization and model optimization methods for model portability (applying a classification model to different images). This work is based on relatively simple (compared to the data models developed for multi-angle LU/LC classification) VHR multispectral classification models and the ability to port these models between images in multi-angle data by correcting the observed data space for distortions caused by physical mechanisms, such as solar illumination and the atmosphere.

When discussing the portability of a classification model, the performance is a combination of the model properties, the physical conditions of the observation environment (including the

atmosphere and illumination conditions), properties of the collection platform, and characteristics of the observed surfaces. It is theoretically possible to improve the model portability problem by addressing distortions caused by all of these effects. However, this would be an exceedingly complex analysis with many sources of variability. Therefore, in the present work, the pursuit of model portability improvements is limited to physically-based removal of distortions due to the mechanisms of the atmosphere, observation platform, and solar illumination in order to analyze the impact of these basic effects. In the presented work, this is referred to as data normalization because techniques, such as atmospheric compensation and band ratios, transpose the satellite data into a normalized regime where, theoretically, data from multiple normalized images can be considered as a single data space.

Techniques to optimize model portability are normally difficult to investigate. This is due to natural changes in real and observed ground reflectivity. Real ground reflectivity can be affected by many things, including aging of the surface material and reflectance changes due to a material's inherent bidirectional reflectance distribution function (BRDF). Observed ground reflectance can be affected by illumination conditions and atmospheric changes when images are collected at different times. In addition, the VHR data space is particularly complex due to the capability of VHR satellites to collect images at high-off nadir angles with associated solar illumination and observed atmospheric effects. This is further complicated by the temporal distribution (and accompanying seasonal and atmospheric changes) of multiple VHR images.

In-track multi-angle data, such as that provided by highly agile satellites like DigitalGlobe WorldView-1 and WorldView-2, provides images of a single scene, collected from different observation angles, during a very short period of time. This creates a sequence of images with relatively static atmospheric and illumination conditions. With this data, the only changes in the scene are due to observation angle and surface reflectance properties. This creates an opportunity to study the model portability problem absent the temporal variation, changing illumination conditions, varying surface reflectance, and changes in atmospheric composition observed in multiple images of a single scene taken at different times. While the data space is still complex, the elimination

of these effects can provide a unique insight into data space normalization techniques that can improve the portability of a machine learning LU/LC model.

The results from the simplified test case of the in-track model portability experiments are then applied to the composite multi-angle data set. This data set is compiled from a multi-temporal sequence of VHR imagery with each image collected at a different view-angle. This provides a similar data set over a single target as that of the in-track sequence. However, the data space is more realistic with the same noise sources that are normally present in multiple VHR images (see Chapter 5 for a deeper discussion of the data space distortion sources). This provides the opportunity to directly measure the portability of a classification model, in normalized data spaces, with a realistic data set that spans the observation space of single image collections.

1.3   Satellite Multi-Angle Data Types

In this dissertation, the problems of LU/LC classification and LU/LC model portability are addressed using the unique capabilities of multi-angle satellite data. These capabilities, including the ability to measure the angular dependence of surface reflectance and the ability to densely sample the angular data space, are leveraged to create experiments that explore methods to improve land-use/land-cover classification of satellite imagery.

Historically, satellite-observed multi-angle surface effects have been explored using low to moderate resolution Earth observation satellites such as AVHRR (Advanced Very High Resolution Radiometer), Landsat, and MODIS (Moderate Resolution Imaging Spectroradiometer). The data collected by these satellites are typified by:

- Moderate spatial resolution: Ground sample distance on the order of 100 m to 1 km.

- Nadir observation: Data is collected at and around nadir as determined by the swath width.

- Wide swath widths: Wide-angle observations with image widths on the order of 2500 km.

- Multispectral sampling: Simultaneous observation of dozens of spectral bands throughout the optical and infrared regions of the electromagnetic spectrum.

• Frequent surface repeat coverage: One or two day revisit rates for most of the Earth's surface due to the large swath width.

Multi-angle information can be utilized from the data produced by this class of satellite in two ways. It can be extracted from multiple nadir observations, using the intrinsic angularity of the observed swath width, or an instrument that employs multiple camera to simultaneously record observations from multiple angles can be employed.

In the first case, it is possible to build up multi-angle data from nadir observing satellites by accumulating multiple passes of a large swath width instrument. The frequent revisit rate (the time delay between observations for any point on the Earth's surface) creates multiple observations for a given target area at a different location in the swath, and, therefore, at a different angle. This mechanism is illustrated in Figure 1.1.

A single satellite system has been deployed which uses the second arrangement of multiple cameras. The Multi-angle Imaging SpectroRadiometer (MISR) employs nine cameras at $0°$, $26°$, $45°$, $60°$, and $71°$forward and backward along the orbital track to record the reflected multi-angle radiation of the Earth's surface across a 360 km swath width [32]. Data is collected in four optical bands with a spatial resolution ranging from 275 m to 1.1 km. The MISR camera configuration is illustrated in Figure 1.2.

Since this is the first, and only[a] , satellite purpose built for multi-angle observation, the benefits and limitations of this type of data have been studied extensively [64, 68, 100]. From this research, many products have been developed that use the multi-angle information of the nine camera array [16, 65, 108]. However, the system is limited by the relatively low spatial resolution of the cameras in which any recorded pixel is a mixture of real land-cover classes.

To resolve the mixed-pixel limitation, very-high spatial resolution satellite data must be used. This imagery is typified by spatial resolution on the order of 1 meter or less, providing observations with pure pixels that more accurately represent the spectral signature of land-cover classes. With

---

[a] The Japanese instrument ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer - on the NASA Terra satellite has a single-band dual-angle multi-camera system that is most often used for the derivation of digital surface models.
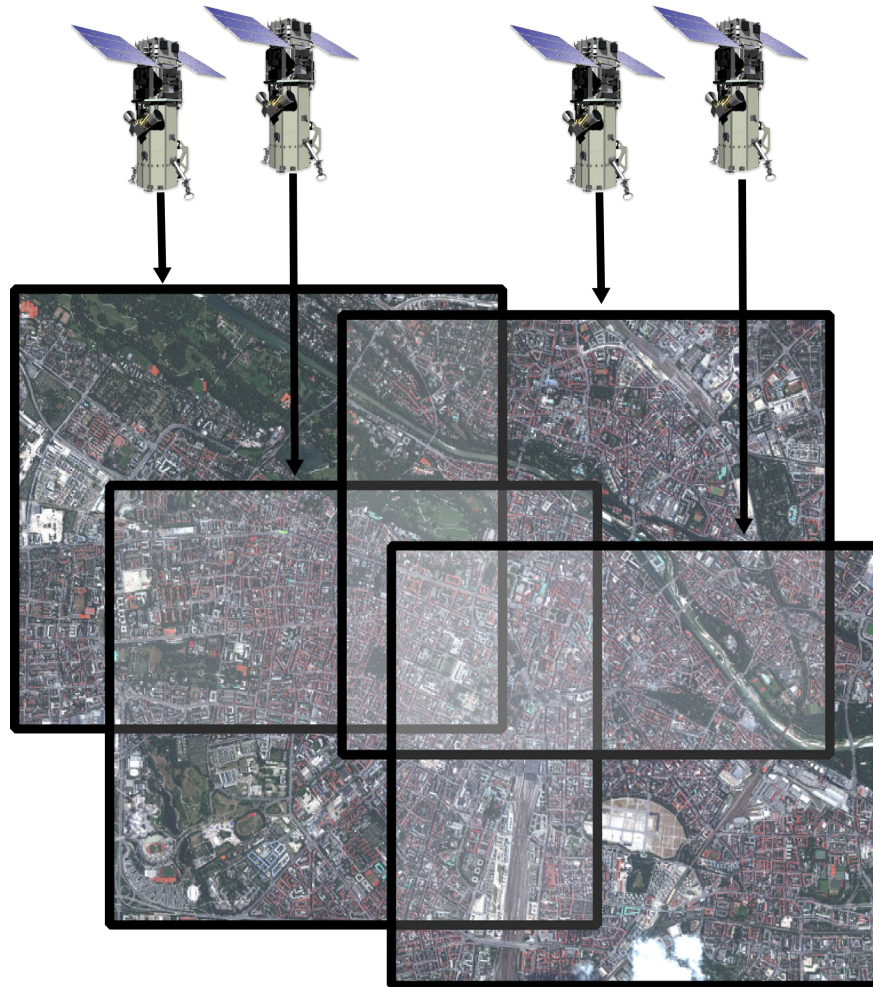
Figure 1.1: An illustration of multi-pass multi-angle data from nadir observing satellites.

this spatial resolution, sufficient to distinguish typical man-made and urban features on the Earth's surface, the observed multi-angle data fundamentally contains different information and opens new avenues of study.

Publicly available data of this type is relatively new, with the first satellites becoming operational around the year 2000 [47]. With the successful launches of the IKONOS (GeoEye, 1999), EROS-A1 (ImageSat, 2000), and Quickbird (DigitalGlobe, 2001) satellites, commercial companies began publicly offering VHR imagery.
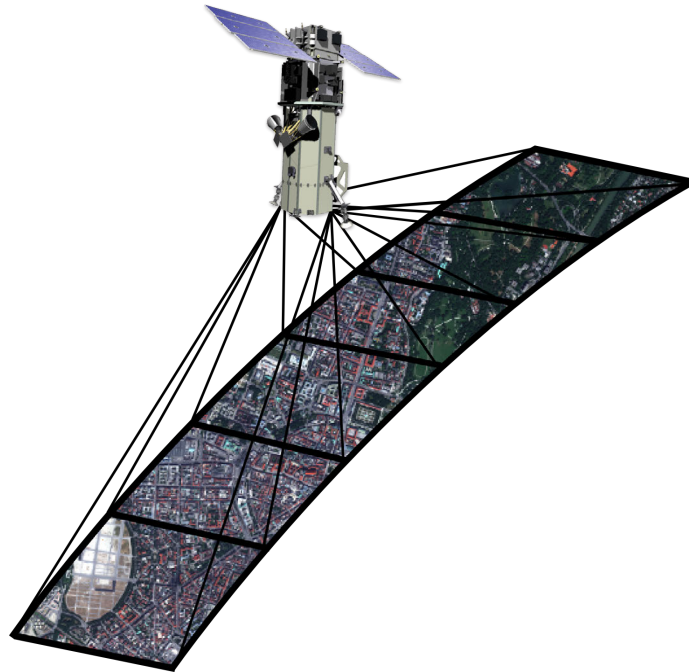
Figure 1.2: An illustration of multi-camera multi-angle data.

However, these satellites have operational capabilities significantly different than the large-area Earth observation platforms described above. This is due to different performance and instrumentation requirements (sensor size, sensitivity, observation angle, slew speed, altitude, etc.) associated with acquiring very-high spatial resolution imagery. The general performance parameters are:

- Very-high spatial resolution: Ground sample distance on the order of 1 m.

- Steerable observation angle: Data is collected by slewing the satellite to the off-nadir angle required to observe a target.

- Narrow swath widths: Full swath width on the order of 15 km.

- Multispectral sampling: Multispectral (often limited to four bands) in the optical region of the electromagnetic spectrum.[b]

---

[b] The next-generation satellite WorldView-3 is slated to be the first VHR system to employ 28 "super-spectral" bands in both the visible, near infrared, and short wave infrared electromagnetic regions. [2]

- Limited repeat coverage: Theoretical revisit rates of one to four days, with real revisit rates dependent on commercial and operational constraints.

The second generation of commercial VHR satellites are now becoming operational. These satellites generally have three additional capabilities:

- Increased collection speed through the inclusion of high-speed control systems such as control moment gyros (CMGs).

- Improved spectral observations providing spectral bands beyond the red, green, blue, and near-infrared collected by the first VHR satellites.

- Additive collection capacity augmenting the current fleet of satellite systems and increasing the actual repeat imaging of a given location.[c]

In this dissertation, data collected by the second generation WorldView-2 satellite is used to create two different multi-angle data sets: in-track multi-angle data and composite multi-angle data.

The improved speed of the WorldView-2 platform by the CMGs provide the opportunity to create dense in-track sequentially-collected imagery. This is illustrated in Figure 1.3. Previous satellite systems, if they were capable of off-nadir data collection, were only able to collect four to five images during a single overflight due to the limited slew speed of the satellite. However, with CMGs, the WorldView-1 and WorldView-2 platforms have 10x better acceleration which reduces the time necessary to slew the satellite over 200 km of the Earth's surface from approximately 50 seconds to 7 seconds. This increased speed directly translates to a dense observation of the angular data space over a target. However, due to the operational requirements of the WorldView-2 platform, in-track multi-angle data is relatively rare. It is discussed in this dissertation as an exploration of future operational possibilities and as a simplified test case of the more common, but significantly more complicated, composite multi-angle data.

---

[c] This would be true of additional first generation satellites that were launched regardless of CMGs or increased spectral performance.
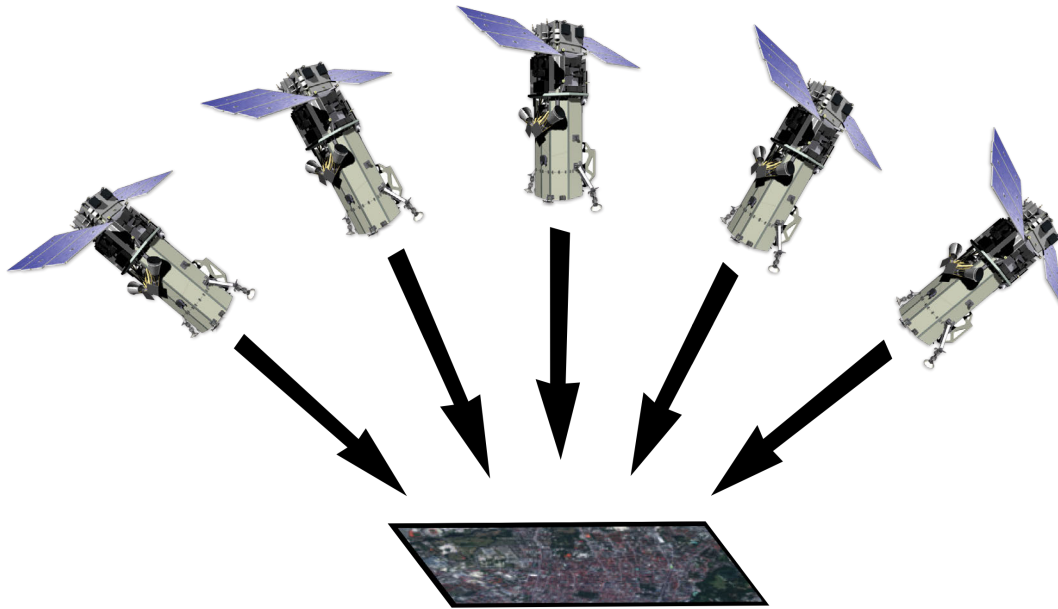
Figure 1.3: An illustration of in-track multi-angle data from a highly agile satellite.

Composite multi-angle data sets are composed of images collected over a number of orbits, usually as collections of opportunity, over the course of a number of months or years. This creates a data set with a variety of observation angles collected at various time throughout the year. This is illustrated in Figure 1.4. This type of data introduces the additional complexities of atmospheric interference, seasonality effects, and varying collection geometries when compared to the in-track multi-angle case. However, compared to the limited availability of in-track multi-angle data, composite multi-angle data sets are more common. This is especially the case in urban areas that are of particular interest to commercial satellite operators, and are, therefore, frequently observed by VHR satellite systems.
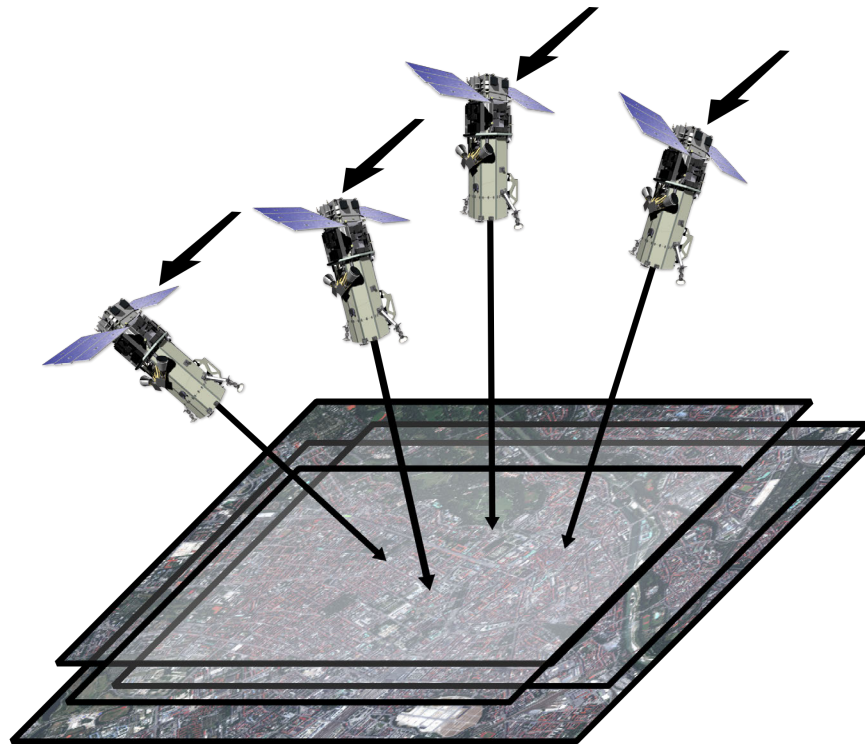
Figure 1.4: An illustration of composite multi-angle data built up from multiple satellite passes.

1.4   Project Overview

In this dissertation, additional information is provided for the urban classification problem through the use of multi-angle satellite observations. These observations can take the form of both temporally dense single pass multi-angle collections, as enabled by the DigitalGlobe satellites WorldView-1 and WorldView-2, and multi-temporal multi-angle collections, as are common over urban areas repeatedly observed by commercial satellites.

This additional knowledge of the urban scene provided by the multi-angle data creates the ability to both improve classification performance and include urban land-use classes, such as bridges and high-volume highways, which are normally difficult to classify in multispectral urban data. Optimized methods that improve the utility of the information as well as difficulties that arise when working with this type of data are explored. The usefulness of the information for urban

classification is also discussed using quantitative classification performance and feature extraction measurements.

This dissertation describes two ways in which multi-angle information can contribute to improved land-cover classification:

- Data Fusion/Feature Extraction: Improving the accuracy of urban classification by extracting multi-angle and multi-temporal information from multi-angle satellite data sets.

- Model Portability: Improving the ability automatically apply multispectral land-cover classification models to a large area by testing data space normalization techniques on multi-angle data sets that simplify the large-area model portability problem.

Within these two goals, the primary objective of this project can be broken down into five main areas of research:

(1) Determine improvement in standard classification accuracy available by including multi-angle information for standard urban land-use classes.

(2) Identify novel classes that can be classified using multi-angle data that are otherwise not extracted by typical urban classification schemes (bridges, skyscrapers, etc.)

(3) Demonstrate that the information used to improve the classification accuracy and increase the number of classes is available in both in-track multi-angle data sequences and more frequently available composite multi-angle data.

(4) Measure the classification model portability using data space normalization techniques using the simplified test case of an in-track multi-angle sequence.

(5) Measure the classification model portability using data space normalization techniques in the more complex data space of the composite multi-angle sequence.

This research is divided into five additional chapters. Chapter 2 presents a literature review describing the current state-of-the-art of relevant multi-angle classification and model portability

techniques. Chapter 3 describes the two multi-angle data sets (in-track and composite) used in these experiments. This is followed by a short introduction in Chapter 4 of the experiments carried out as part of this dissertation. Chapters 5 and 6 describe the experiments, methods, and results of the model portability and classification projects, respectively. This is concluded by a summation of the results, lessons learned, and recommend future research directions in Chapter 7.

Chapter 2

Related Background Studies

2.1 Multi-Angle Data

This section provides a brief introduction to multi-angle surface data followed by a description of the current state-of-the-art in collection and usage of multi-angle data from both aerial and satellite platforms.

2.1.1 Multi-Angle Surface Data

In its most basic form, observed optical light reflected from a surface depends on the texture and reflectance properties of that surface. The reflectance determines how much of the incoming light is absorbed (vs reflected) and the texture determines the direction that the reflected light is scattered. This can be a very complicated problem [45, 52], but for the practical circumstances encountered in most optical remote sensing, these properties can be described relatively directly.

The amount of reflected energy can be described as the simple difference: $reflected = incident - absorbed$. Therefore, reflectance, as is usually used to describe a surface is $reflectance = reflected/incident$. The physics that determine how much of incoming light is absorbed is beyond the scope of this description, but working "ground truth" values used in remote sensing are typically determined experimentally [36].

The remotely observed reflectance also has an angular dependence for all real materials. The percent reflected light that is observed at a particular observation and illumination angular is described by the materials bi-directional reflectance distribution function (BRDF). This effect has

been described extensively in literature, but is essentially a method of encompassing the variation in reflectance from a surface due to both the observation angle and illumination angle [94].

Many attempts have been made to characterize surface BRDF properties with in-situ measurements [11, 89]. These generally implement multi-angle instrumentation from which a full hemisphere of information can be acquired at a single solar illumination angle (a short period of time observing a real surface) by moving an observation arm or instrumenting a full hemisphere measurement. This sort of system can either employ a spectrometer to measure a fine spectral resolution of the reflected solar radiation or can describe small variations in the spectral signature over the observed angles by illuminating the surface with a known light source.

Using this dense spectral and angular information, the observable surface properties of different materials can be well characterized. This characterization can then be used to identify surface materials from remote sensing observations. However, at the practical level of remote sensing observation, the measured properties are often a superposition of several surface types per observation. In the case of low or moderate resolution imagery (100's of meters or more), this combination of properties takes place over a very large area, leaving the accumulated spectral signature with possibly dozens of surface types integrated into a single recorded observation. Therefore, the moderate resolution satellite studies that utilize multi-angle data are limited to relatively broad (compared to the VHR studies presented in this dissertation) surface descriptions [35, 87]

With this problem posed, it is clear that the finer spatial and spectral, and by extension angular, resolutions can fundamentally reduce the complexity of the posed multi-angle problem. Applying higher resolution spatial and spectral observations to multi-angle observations provides an additional data dimension that can be leveraged for information retrieval and, therefore, improved segmentation of the data space. If the complexities associated with this additional information are appropriately overcome, the result is improved identification of individual objects, a more accurate representation of the angular data space, and improved classification accuracy for surfaces that are otherwise difficult to identify in satellite imagery.

2.1.2   Multi-Angle Aerial Data

In addition to in-situ multi-angle measurements, there are two additional platforms commonly used to collect multi-angle data: aerial and satellite. However, unlike in-situ measurements that use surface instruments to measure a full hemisphere of information very quickly, multi-angle aerial and satellite observations are generally individual measurements that must be accumulated over some period of time to build up the multi-angle information. It is this requirement that is both the power and drawback of remote multi-angle observations. With these built up observations, multi-angle information can be acquired over a large area, but the methods for leveraging this information require dealing with noise sources associated with atmospheric, illumination, and real surface changes between individual observations.

In many ways, these aerial and satellite platforms provide very similar data at the information level. However, there are significant operational and engineering differences that generally afford aerial data with higher spatial and spectral resolutions.

Previously, aerial platforms that collected multi-angle information were used to show that the additional angular information contains unique descriptions of otherwise similar surface cover [8]. More recently, aerial measurement of surface multi-angle properties have become more common due to competition in on-line mapping products. A key ingredient to these mapping platforms is aerial data - due to the high spatial resolution afforded by the relatively close observations and current regulatory limitation on satellite spatial resolution. This data is often collected with significant overlap due to the mapping advantages afforded by multiple views of a given location and the desire to develop three-dimensional models [86]. Additionally, this overlap has been leveraged to extract multi-angle information for the purposes of surface classification [54].

2.1.3   Multi-Angle Satellite Data

Multi-angle satellite imagery has previously been explored in two general forms. The first leverages the swath width of an instrument and multiple overflights of a single instrument to build

up different look angles of a target. This approach has been explored extensively with moderate and low spatial resolution imagery [88, 99]. The second approach uses either multiple cameras or pointing capabilities to generate multiple views of a single target during a single overflight.

Two previous satellite missions were designed to collect multi-angle data with this second approach. The first mission, the Multi-angle Imaging SpectroRadiometer (MISR), was launched by NASA in 1999 and uses nine cameras pointed up to 70.5°off-nadir to collect imagery with a spatial resolution of 275 m to 1.1 km (dependent on camera viewing geometry) in four optical bands [32]. The second satellite, the Compact High Resolution Imaging Spectrometer (CHRIS), was launched by the European Space Agency (ESA) in 2001 and uses pointing of the instrument to collect up to five images of a given target at up to 55°off-nadir with a spatial resolution ranging from 17 m and 34 m at nadir with up to 62 optical spectral channels (dependent on collection mode) [9, 27].[a]

The DigitalGlobe WorldView-1 and WorldView-2 extends this second approach with high slew rate instrumentation to allow for collection of very high resolution multispectral imagery at dozens of view angles over a given target. This type of data has recently been shown to be useful for several areas of research including: bathymetry [60], surface model extraction [26], and land-cover classification [66].

This dissertation also discusses the capabilities that are available through the use of composite multi-angle data. This data is composed of multiple collects from pointing satellites, such as WorldView-2. This data can be thought of as a collection of multi-temporal data. In fact, various aspects of the multi-temporal characteristics are leveraged in these discussions. However, the unique aspects of this work are in the extraction and usage of the multi-angle information present in such a dataset.

---

[a] A third satellite mission, Advanced Spaceborne Thermal Emission and Reflection Radiometer - ASTER - currently operating on the NASA Terra satellite, has a single optical channel the collects off-nadir data for purposes of elevation model generation.

2.2   LU/LC Machine Learning Model Portability

Model portability refers to the ability to move a data model, describing the relationship of remote sensing observations to a measurement of interest, from one observation to another. This is of particular concern for applications of VHR imagery. While this imagery provides superior spatial resolution and creates the opportunity to identify individual surface objects, it also comes at the operational trade-off of swath width (Section 1.3). Therefore, in order to accurately create a LU/LC map over a large area, a model must be applicable over multiple observations.

Techniques to train a land-cover classification model and accurately port the classifier from one image to another are currently not well understood because the complexities (spatial, spectral, etc.) of VHR imagery generally prevent a model developed for one image from being applied to another. In order to develop this capability, one of two corrections must be made to the model/data system:

(1) Transfer Learning: The classification model must be dynamic and flexible enough to accurately understand, and adapt to, the distortions between the multiple images [83, 105].

(2) Domain Adaptation: The data from each image must be explicitly converted into the data space from which the model was trained [49].

The first approach has been extensively studied. The basic approach is to take a model and apply it to different, but related, data sets by adjusting the model for the new data [83]. One example would be a statistical classifier that is adjusted for variation between two images due to changes in atmospheric conditions by adjusting classification coefficients for all spectral bands according to the observed radiance in a known dark object.

The second approach, domain adaptation, is an active area of research and no individual method has gained general acceptance. Most of the recent efforts have attempted to mold the data manifolds based on statistical comparisons and manipulations between the data spaces of interest. [103]. Other attempts to normalize the data space through the use of contextual information

showed that the portability of a classification model can be improved through the use of texture filters [30].

In this project, domain adaptation is pursued through the use of physically based data space normalization and explicitly modeling distortion effects, such as atmospheric compensation and illumination correction. The ability to move an arbitrary image into a standard data space through physical modeling of the distortion effects is explored by directly measuring the portability of a LU/LC classification model in multi-angle data.

2.3   Multi-Angle Derived Surface Models

The use of optical imagery for the extraction of digital surface models (DSM) and digital elevation models (DEM) has been actively pursued since the Earth Terrain Camera was flown aboard Skylab in 1973/74 [102]. The approach generally uses scene changes from multiple view angles to extract elevation. This approach is similar to the way humans view three dimensional scenes. The difference is that DSM algorithms attempt to use mathematical constructs to simulate the human correlation capability [59, 75]. The majority of DSM algorithms implement some version of cross correlation, using geometric knowledge to extract the distance of each observed and correlated pixel.

More recent DSM extraction algorithms have used a multi-tiered matching approach, matching not only with the spatial correlation, but utilizing area matching and edge detection. Additionally, some algorithms, such as NGATE, use multi-image matching whereby multiple image pairs are used to validate the optimal correlation. In this way, the DSM extraction algorithm is able to leverage multi-angle collections, that are the subject of these studies, to improve the DSM accuracy.

2.4   LU/LC Classification with Height Information

The application of elevation data to image classification has been explored since the 1980's. The majority of these early publications, however, took the form of geographic studies, including water shed analysis [7, 114] and terrain classification [48, 77]. These studies focused on the utility

of low resolution stereo imagery to study elevation information available from extracted surface models.

Several studies have shown the utility of elevation and height information for the development urban LU/LC classification maps. Lemenosu *et al.* [61] used the relatively coarse surface and elevation data available from the Shuttle Radar Topography Mission digital elevation model (SRTM-DEM) and National Elevation Dataset (NED) to derive the average height above the surface for two North American urban areas. Sportouche *et al.* [98] showed the utility of multi-source SAR/optical data fusion for building extraction by using SAR data to both refine a coarse optical-based building extraction and to perform building height retrieval. Haitao *et al.* [40] presented results showing that the fusion of a LIDAR based digital surface model (DSM) and aerial optical imagery with a support vector machine can provide excellent classification accuracy in an urban test case.

Previous single-source height classification studies utilized SAR or LIDAR data to extract height information for urban classification. For example, Haala *et al.* [39] demonstrated the use of LIDAR to extract buildings and trees in an urban area. Song *et al.* [97] showed the utility of LIDAR intensity data for urban classification and concluded that height information would resolve the least separable classes of similar intensity. Antonarakis *et al.* [6] continued this approach by including not only height, but also the LIDAR point distribution frequency to enhance the classification scheme. Tison *et al.* [101] have shown the use of simultaneous classification and height extraction of an urban area using SAR data.

2.5   LU/LC Multi-Angle Classification

The use of the multi-angle imagery has been explored extensively in the field of Earth science. By far the most extensive use of multi-angle imagery in scientific research has been for the calculation of atmospheric aerosol [43,51,65,71]. However, the use of multi-angle imagery for the purpose of LU/LC classification also has significant literature background. Most applications, though, focused on moderate resolution vegetation studies [64,100] (using moderate resolution multi-angle informa-

tion described in Section 1.3) due to the available spatial resolution and the interest in the utility of bidirectional reflectance distribution function (BRDF) information for classifying vegetation.

Some previous multi-angle classification work was performed at higher spatial resolution through the use of unmanned aerial vehicles [3], though this is also generally rural work. The CHRIS instrument, in high spatial resolution mode (17 m), has provided some opportunity to explore urban classification. This work has focused on mixed use areas and the differentiation of urban features from rural classes with multi-angle multi-temporal data [29, 33].

The current work extends these previous studies to include a more dense sampling of the multi-angle space and explores the application of multi-angle data at very high spatial resolutions. With this information, in addition to the additional spectral bands provided by WorldView-2[b] , LU/LC classification is improved through the extraction of multi-angle BRDF information not present in single image data.

---

[b] Relative to the four optical bands typically observed by VHR satellites

Chapter 3

Data

The experiments carried out as part of this dissertation use two distinct multi-angle data sets - in-track multi-angle and composite multi-angle. The in-track multi-angle data set consists of imagery that is collected sequentially during a single over-flight of a target area using the fast retargeting capability of the satellite to repeatedly image the same target. In contrast, composite multi-angle data sets are created from a collection of opportunistic satellite observations, individually collected over a period of months or years, that are available for a target area.

The in-track data is dense in both time and angle-sampling along the satellite track, but this type of data set is relatively rare. An in-track collection requires a high slew rate satellite and a significant portion of the satellite's resources during an orbit. For these reasons, there are very few of these data sets available. Most were taken during the testing and evaluation period of WorldView-2, the second DigitalGlobe satellite, in addition to WorldView-1, capable of the type of in-track rapid retargeting required for data collections of this type. On the other hand, composite multi-angle data sets are collected over a number of orbits with the angular and temporal sampling determined by the available images for a given target area. Because this type of data set leverages previously collected images, it is more common in urban areas repeatedly imaged by commercial satellite systems.

Due to the resolution and performance requirements related to the resolving urban features and collecting multi-angle observations, the data utilized here will focus on data from the Digital-Globe satellite WorldView-2. WorldView-2 collects eight bands of spectral data ranging from 400
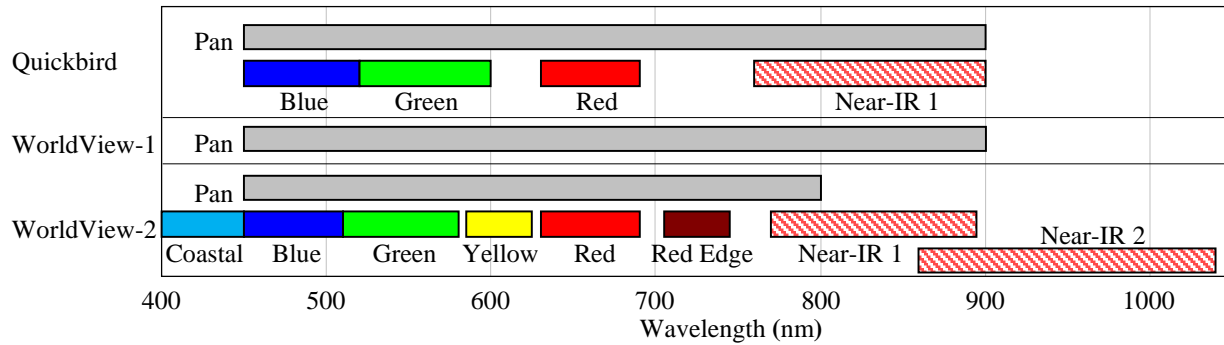
Figure 3.1: Observed optical bands of the three DigitalGlobe satellites.
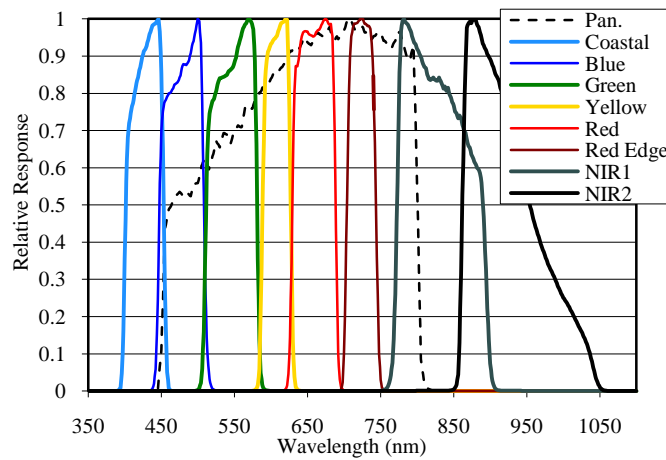
Figure 3.2: Spectral response of the WorldView-2 panchromatic and multispectral bands. The panchromatic band is sensitive to radiation from approximately 450 nm to 800 nm. The full multispectral range extends beyond this limit to include one shorter wavelength band sensitive to radiation down to approximately 400 nm and two bands in the near infrared (NIR) that extend as far as 1050 nm.

nm to 1050 nm with a spatial resolution of 2 m. In addition, it simultaneously collects panchromatic imagery with 0.5 m spatial resolution over the range of 450 nm to 800 nm. Figure 3.2 shows the spectral response curves of these bands. This spatial resolution and multispectral capability provides improved classification performance with respect to previous four band satellites [69].

The spatial and spectral capabilities of WorldView-2 are complemented by the satellite's agility. The camera has a maximum look angle of approximately 40 degrees off-nadir (higher angles are selectively commandable) and can slew across 300 km of the Earth's surface in 9 seconds. The satellite system can collect up to 1 million km^2 per day, doubling the previous collection

capability of the DigitalGlobe constellation. This unique agility and collection capacity is what enables WorldView-2 to acquire the angularly dense imagery of an in-track multi-angle data set.

In this project, two data sets are studied. One is an in-track multi-angle sequence collected over Atlanta, Georgia (U.S.A.) and one is a composite multi-angle data set created from images collected over Lakewood, CO (U.S.A.). The specifics of the datasets are discussed below.

3.1   In-Track Multi-Angle Data

The WorldView-2 multi-angle collection took place over Atlanta, Georgia, (U.S.A.) in December 2009. The full data set contains 27 images with satellite elevation (angle between the horizon and satellite as viewed from the image target) from 60°looking northward (from the target area) to 25°looking southward with a maximum satellite elevation of 81.5°. Figure 3.3 shows the elevation and azimuth of the 27 images and their distribution along the collection track. The images are generally well spaced with a regular collection timing with the sequence starting relatively late in the orbit and completing near the southern horizon. The images were collected within a two minute time frame with a sun azimuth and elevation of 164°and 31°, respectively.

For the purposes of this study, the analyzed images are limited to a subset of the full sequence that have a moderate (relative to the satellite's capability) off-nadir observation angle of less than 30°. This results in a sequence of 13 images distributed between an elevation of 57°looking northward from the target area to an elevation of 59°looking southward, with a maximum satellite elevation of 81.5°. Figure 3.4 shows the elevation and azimuth of the 13 images and the distribution along the collection track. The images were collected sequentially within a two minute time frame with a sun azimuth and elevation of 164°and 31°, respectively.

The study area is near to, and encompassing part of, downtown Atlanta. This provides an interesting and complex urban scene with many man-made and natural class types. Figure 3.5 shows an image of the area. The scene measures 7628×5064 pixels in the panchromatic band (50 cm spatial resolution) and 1907×1266 in the multispectral bands (2 m spatial resolution). The scene
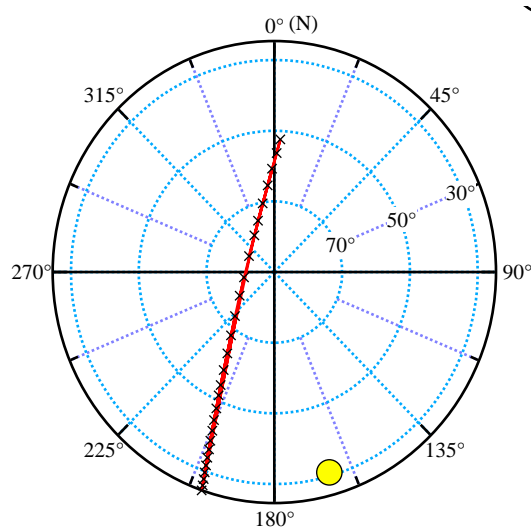
Figure 3.3: Ground observed azimuth and elevation of the satellite for each image observation (black marker) as well as the azimuth and elevation of the sun (yellow circle) during the Atlanta multi-angle sequence. Azimuth is plotted angularly clockwise (north = 0°, east = 90°, south = 180°, west = 270°) and elevation is plotted radially from the center (ground nadir = 90°, ground horizon = 0°).
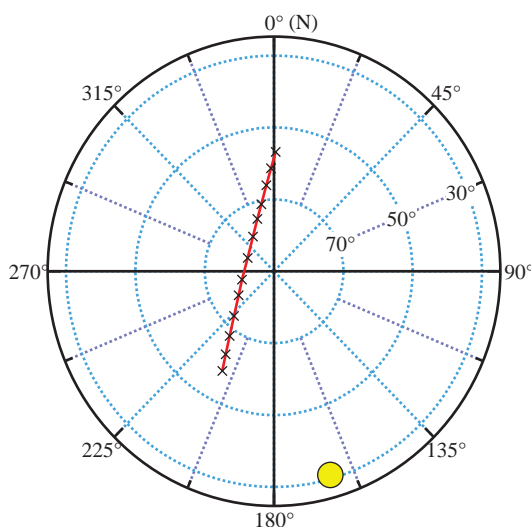
Figure 3.4: Ground observed azimuth and elevation of the satellite for each image observation (black marker) as well as the azimuth and elevation of the sun (yellow circle) during the Atlanta multi-angle sequence. Azimuth is plotted angularly clockwise (north = 0°, east = 90°, south = 180°, west = 270°) and elevation is plotted radially from the center (ground nadir = 90°, ground horizon = 0°).

Figure 3.5: Color infrared image (using the NIR2 band) of the target area near downtown Atlanta, Georgia.

includes a number of large buildings, urban canyons, commercial centers, and suburban residential areas that include a mixture of community parks and private housing.

It is important to note that the image sequence was taken in winter, soon after the launch of WorldView-2. This significantly impacts the vegetation observations in the sequence. A majority of the trees in the area are deciduous and were, therefore, leaf-off at the time. Consequently, some observed spectral signatures are due to the superposition of an underlying class and the leaf-off tree canopy, convoluting the classes in the machine learning system.

Furthermore, the atmospheric effects observed across the multi-angle sequence vary significantly due to the increasing optical path length at lower satellite elevations. In the visible and near-infrared portions of the electromagnetic spectrum, the total radiance at the sensor $(L_{sensor,\lambda})$

can be described as:

$$L_{sensor} = L_{solar} + L_{up} + L_{down} + L_{adj} \tag{3.1}$$

where all parameters are assumed to have a wavelength (λ) dependence. L_{solar} is the solar radiance reflected from the target, L_{up} is the scattered solar radiance from the atmosphere to the sensor, L_{down} is the downward analog of L_{up} reflected from the target, and L_{adj} is radiation reflected by objects other than the target and scattered or reflected into the sensor [107]. For a given scene, the L_{solar}, L_{down}, and L_{adj} terms are dominated by the atmospheric composition and solar illumination of the acquisition. Additionally, the L_{up} component can vary significantly for off-nadir acquisitions because it significantly depends on the geometric atmospheric depth at the collection angle.

Calibration of instrument digital number to surface reflectance may be expressed as [91]:

$$\rho_\lambda = \frac{K\pi(L_{senor,\lambda} - L_{up,\lambda})}{\tau_{up,\lambda}(E_{toa,\lambda}\cos(\theta_z)\tau_{down,\lambda} + E_{down,\lambda})} \tag{3.2}$$

where K is the illumination correction factor [55], $L_{senor,\lambda}$ is the radiance measured by the sensor, $L_{up,\lambda}$ is the radiance scattered by the atmosphere upward, $\tau_{up,\lambda}$ is the atmospheric transmittance from the ground to the TOA, $E_{toa,\lambda}$ is the solar irradiance incident at the TOA, θ_z is the solar zenith angle, $\tau_{down,\lambda}$ is the atmospheric transmittance from the TOA to the ground, and $E_{down,\lambda}$ is the diffuse irradiance at the surface (when the adjacency effect, L_{adj}, is not explicitly included). In this way, surface reflectance can be retrieved with atmospheric, solar, and satellite look-angle parameters appropriate for the a given image.

The upwelling radiance (energy scattered at the sensor by the atmosphere) [94] is a major component of the atmospheric spectral distortion effect. This is shown in Figure 3.6 for the eight WorldView-2 spectral bands over the analyzed observation angles (calculated using MODTRAN [13] for the satellite geometry and atmospheric conditions present during the multi-angle collection). This is driven by Rayleigh scattering and has a λ^{-4} dependence on wavelength, therefore, the effect is more pronounced at lower wavelengths. If this effect is not considered, the imagery will have increased radiance at shorter wavelengths and lower satellite elevation. Consequently, in this study,
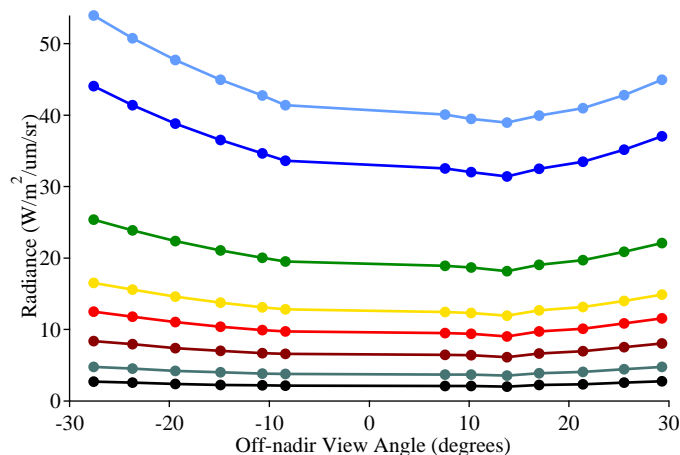
Figure 3.6: Upwelling radiance by spectral band across the in-track multi-angle sequence.

each image is normalized to surface reflectance using FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) [23], to account for view-angle atmospheric effects.

Figure 3.7 shows a practical example of these spectral distortions by comparing the top-of-atmosphere reflectance and atmospherically corrected surface reflectance for the shadow and evergreen classes. These classes are particularly interesting because they have well known spectral signatures. The shadow class should be relatively dark with minimal reflectance across the various angles and the evergreen class should display the well known near-infrared plateau as well as a much smaller green peak in the shorter visible wavelengths. In the shadow class, the scattering of short wavelength radiation clearly dominates the uncorrected reflectance curve. Atmospheric correction produces an appropriately flat signature. The evergreen class is similarly corrected and exhibits the near-infrared plateau as well as a heightened green band in the shorter visible wavelengths. In addition, the red edge band is well positioned to measure the reflectance change between the visible bands and near infrared plateau.

One of the unique aspects of this data set is the ability to directly observe a dense angular sampling of multi-angle reflectance by class. As an illustration of this, Figure 3.8 shows plots of the mean reflectance for eight classes across the multi-angle sequence. These are produced by
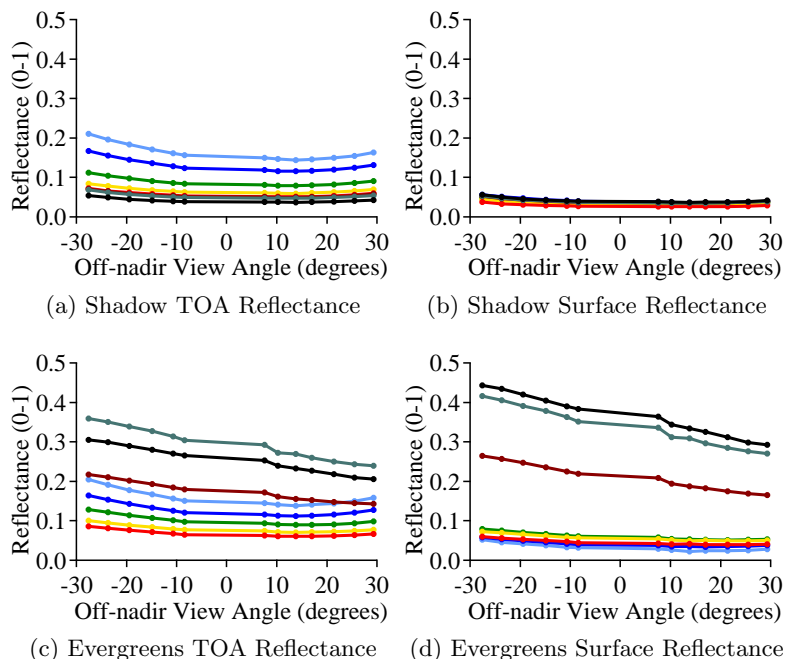
Figure 3.7: Average top-of-atmosphere (3.7a, 3.7c) and surface (3.7b, 3.7d) reflectance of the World-View-2 multispectral bands for the shadow and evergreen ground survey classes with markers indicating the observation angles of the imagery used in this study. Color legend: ▬ = coastal, ▬ = blue, ▬ = green, ▬ = yellow, ▬ = red, ▬ = red edge, ▬ = NIR1, ▬ = NIR2.

aggregating the reflectance values of the ground survey validation set for each class across the multi-angle sequence as described in Section 6.3.1.

The plots in Figure 3.8 show examples of land-cover classes with multi-angle reflectance (MAR) curves that are dominated by the four multi-angle spectral information sources:

1) Partial BRDF

2) Solar cross-section observation

3) Temporal variation

4) Specular reflection invariance

The partial BRDF effect is the most difficult to show and validate because it is non-trivial to ensure differentiation of the partial BRDF signal from atmospheric effects without in-situ concurrent BRDF observations [36,89]. Despite this, one indicator of the atmospheric compensation accuracy,
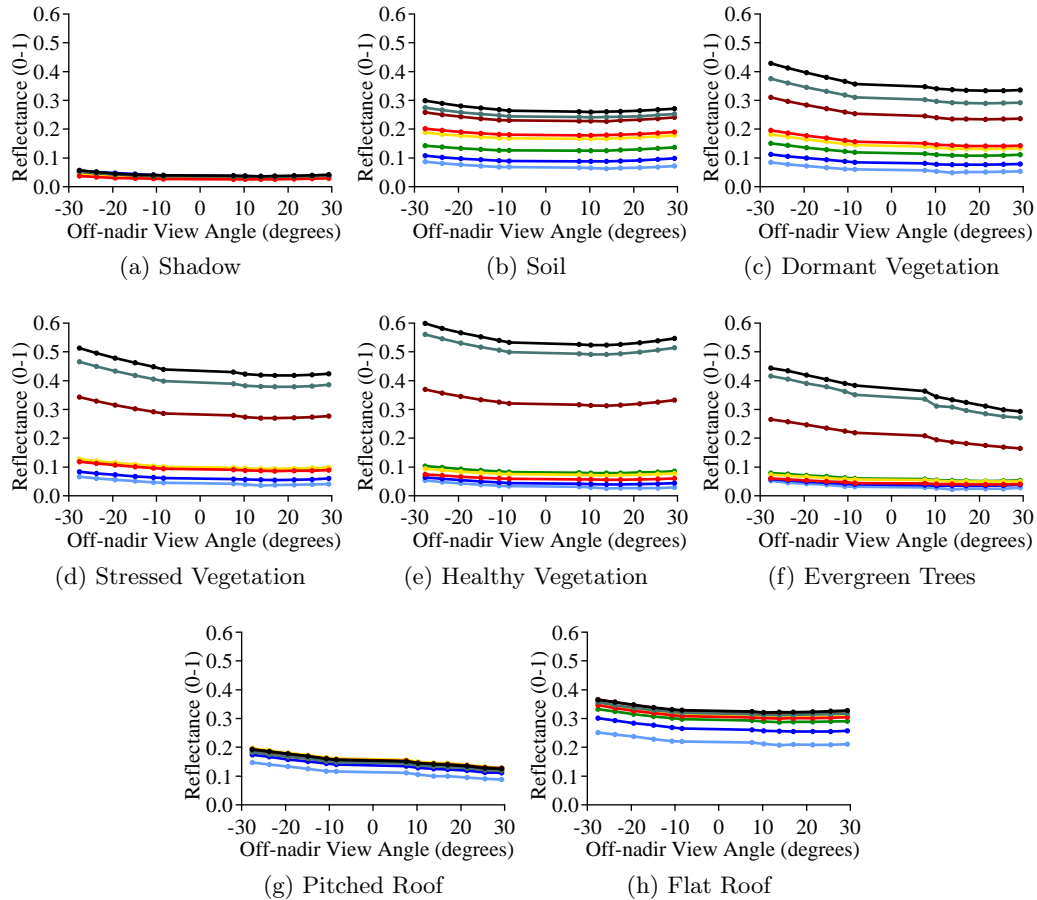
Figure 3.8: Average surface reflectance calculated from the ground survey of the selected classes with markers indicating the observation angles of the imagery used in this study. Color legend: ▢ = coastal, ▢ = blue, ▢ = green, ▢ = yellow, ▢ = red, ▢ = red edge, ▢ = NIR1, ▢ = NIR2.

and, therefore, the accuracy of the remaining BRDF signal, is the near-zero reflectance of the shadow class (Figure 3.8a). The BRDF signal is expected to be present in most vegetation classes [20]. However, the low-lying classes, such as soil (Figure 3.8b), dormant vegetation (Figure 3.8c), stressed vegetation (Figure 3.8d), and healthy vegetation (Figure 3.8e), will be most likely to show this signal because they are not subject to the observed solar illumination cross-section. Indeed, a small spectral variation can be seen after removing atmospheric effects for these classes in the form of a small curvature and/or linear trend across view angle.

The observational solar cross-section dominates the evergreen (Figure 3.8f) and pitched roof (Figure 3.8g) classes. Ideally, in a true-ortho image, the solar observational cross-section would be completely removed at the pixel level. However, this would require an ideal surface model capable of representing subtle changes in elevation within small areas such as pitched roofs and evergreen trees. If this were the case, the only observed effect would be a partial BRDF through the multi-angle sequence. However, the surface model does not generally extract the small subtleties of surfaces a few pixels in size. Therefore, the observed solar cross-section is also generally intact and can be exploited for classification purposes. In the reflectance plots, this effect is observed as a nearly linear slope across observation angle.

Temporal reflectance variation can also provide a signal for certain class types. In these cases, the variation in observed reflectance of randomly oriented vehicle surfaces or movement of vehicles during the multi-angle collection causes the pixel multi-angle reflectance (MAR) to vary significantly. These variations can be modeled by calculating the standard error of the reflectance trend across view angle. Pixels with a large standard error can be used to map areas of high variability in the scene.

The fourth information source, the multi-angle invariance to specular reflection, is not a large effect at the low solar angle of this collection and is not shown.

3.2   Composite Multi-Angle Data

Composite multi-angle image sets are particularly interesting because, despite the seasonality issues discussed earlier, they contain a large amount of information and, due to the prevalence of commercial very-high spatial resolution imagery, can be collected for many cities around the world. Because of this prevalence, any additional urban classification leverage provided by the information in this data set could be directly applicable to other urban projects.

The composite multi-angle sequence used in these experiments was collected over Lakewood, CO (U.S.A.). The image sequence is comprised of WorldView-2 images collected from October 2010 to June 2011. The data set includes a single in-track triplet similar to, but with fewer images

2010												2011					
1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6
◆			◆	◆	◆	◆	◆	◆	◆	◆	◆		◆	◆			◆

Table 3.1: Temporal distribution of WorldView-2 collections in the composite multi-angle data set over Lakewood, CO (U.S.A.). The image sequence consists of 28 images, collected on 26 separate days over a two year period.
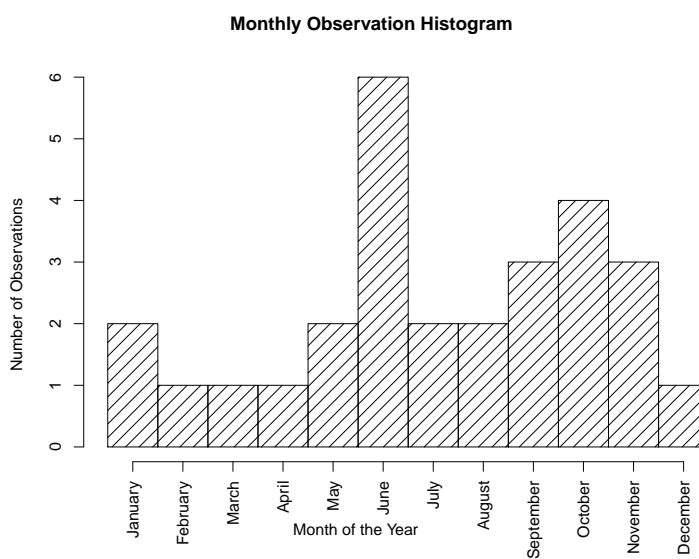
Figure 3.9: Histogram of the Lakewood composite multi-angle sequence monthly image distribution.

than, the data used in the in-track multi-angle classification experiments of Sections 6.2 and 6.3. The months with imagery collected over this two year time period are shown in Table 3.1.

A histogram of the monthly distribution is shown in Figure 3.9. From this histogram, we can see that, the data set is reasonable well sampled through the year and includes images from Winter, Spring, Summer, and Fall. There are generally fewer winter images in the sequence as compared to the other seasons due to frequent presence of snow in the target area.

The composite multi-angle data set is also well distributed across the possible collection geometries. This is illustrated in Figure 3.10. Since the data set is mostly comprised of single images, and most single images are collected near nadir, the majority of the images have a satellite elevation of greater than $70°$.
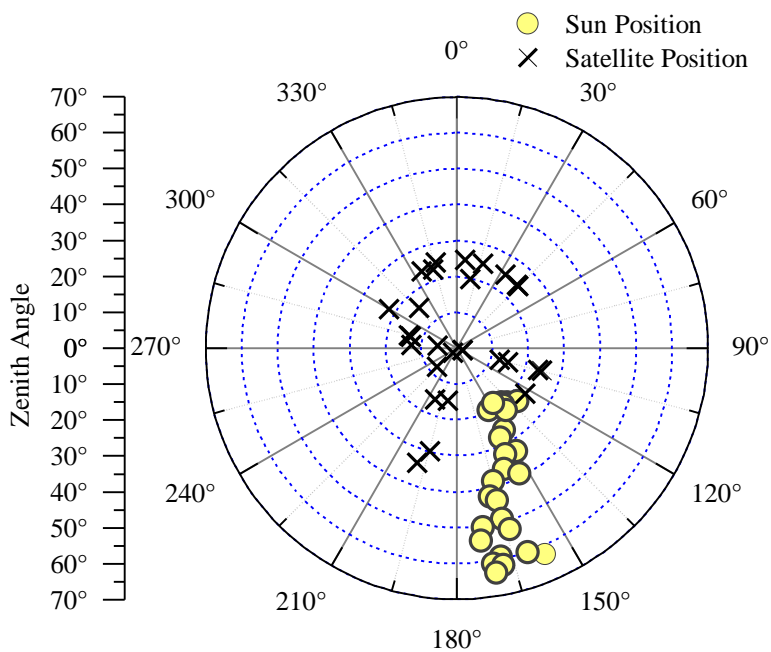
Figure 3.10: Ground observed azimuth and elevation of the satellite for each image observation (black marker) of the Lakewood data set. Azimuth is plotted angularly clockwise (north = 0°, east = 90°, south = 180°, west = 270°) and elevation is plotted radially from the center (ground nadir = 90°, ground horizon = 0°).

The Lakewood, CO target area was chosen due to the frequent WorldView-2 collections near the DigitalGlobe headquarters for testing and calibration purposes. The area is a mixed suburban community that includes residential neighborhoods, commercial areas, major highways, residential roadways, parks, and a series of small lakes. A color infrared image of the area is shown in Figure 3.11.
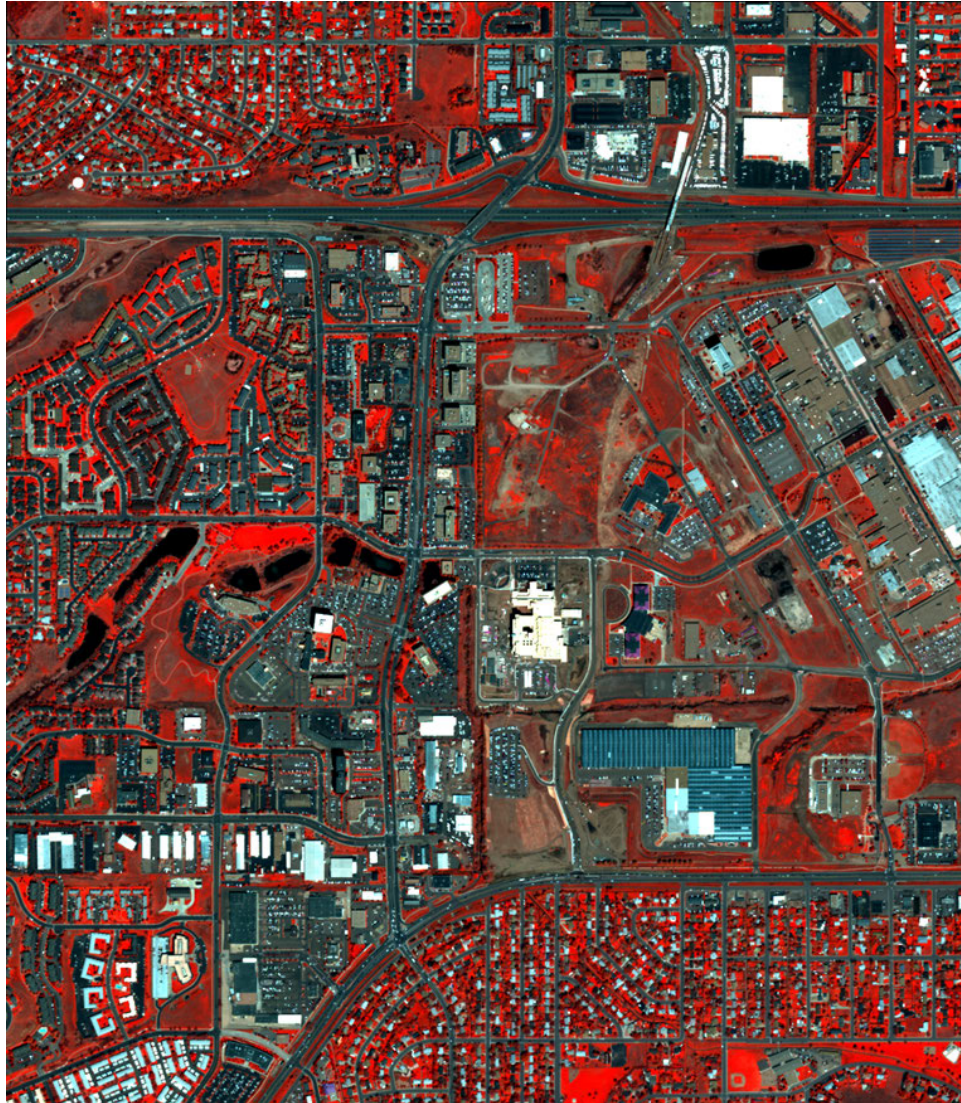
Figure 3.11: Color infrared image (using the NIR2 band) of the target area in Lakewood, CO.

Chapter 4

Overview of Multi-Angle Experiments

This dissertation discusses four related projects that explore two areas of LU/LC classification necessary to create large-area LU/LC classification maps. These projects are derived from the exploration of two LU/LC classification problems applied to two types of multi-angle data (see Table 4.1). The two classification problems are multispectral LU/LC model portability and LU/LC classification using multi-angle derived information; each are introduced in Chapter 1 and described in greater detail in Sections 5.1 and 6.1. Both of these projects are then explored using two types of multi-angle data: in-track multi-angle and composite multi-angle. These data are discussed in detail in Sections 3.1 and 3.2.

In these experiments, no attempt is made to improve classification accuracy from the algorithm perspective. Many previous studies have focus on methods for improving classification performance through the use of optimized machine learning algorithms as briefly discussed in

	Model Portability	Multi-Angle Classification
In-Track Multi-Angle Data	Section 5.2 Model Portability: In-track Multi-Angle	Section 6.2 & 6.3 Classification: In-track Multi-Angle
Composite Multi-Angle Data	Section 5.3 Model Portability: Composite Multi-Angle	Section 6.4 Classification: Composite Multi-Angle

Table 4.1: The four projects presented in this dissertation are an exploration of two types of VHR multi-angle data as applied to multispectral LU/LC classification model portability and improved LU/LC classification though the incorporation of multi-angle information.

Section 1.1. However, the goal of the present study is to determine the degree to which classification performance can be improved through the use of multi-angle data. Therefore, the presented experiments use the same classification algorithm so that the impact of the various multi-angle experiments can be determined separate from the impact of the classification techniques.

The experiments in this dissertation are carried out with the Random Forest algorithm trained using supervised classification data. The Random Forest algorithm [17] is a well known ensemble method that has proven effective and computationally efficient [44] and has been successfully applied to previous remote sensing problems [37]. In all cases, a single configuration of the Random Forest (depth and density) is used for each set of experiments in order to directly compare the classification performance of individual experiments. More on the parameter choices of the Random Forest algorithm is discussed in Appendix A.

Additionally, a feature contribution method is implemented that uses the information gain ratio [90] to extract multi-angle feature contribution by class (further discussed in Sections 6.2 and 6.3). This provides the opportunity to explore the utility of multi-angle data for specific classes as well as for overall classification performance.

Each of these projects is designed to explore a specific aspect of multi-angle data as applied to land-use/land-cover classification. An introduction as well as the methodology and results of the four projects are discussed in detail in Chapters 5 (Multispectral Model Portability) and 6 (Multi-Angle LU/LC Classification).

Chapter 5

Multispectral Model Portability

This chapter details both the in-track and composite multi-angle model portability experiments. These experiments explore the ability to apply a multispectral LU/LC classification model to many different images. While the full model portability problem is exceedingly complex (see discussion below in Section 5.1), these experiments explore the ability to mitigate these complicating factors in order to develop an understanding of effects that determine the portability of a multispectral LU/LC model.

Section 5.1 introduces the model portability problem while Sections 5.2 and 5.3 detail the experimental methods and results of two parallel model portability studies that explore the in-track and composite multi-angle data spaces. The multi-angle experiments are continued in Chapter 6 with the exploration of the multi-angle classification. The results from both the model portability and classification experiments are concluded in Chapter 7.

5.1   Intro: Multispectral Model Portability

The increasing availability of very-high spatial resolution (VHR) imagery provides the opportunity to leverage VHR image libraries to create large scale, very-high spatial resolution LU/LC classification maps. These maps would be unique among currently available products in that they would feature surface material and/or use information with the resolution necessary to describe typical urban features. However, due to the swath width constraints of VHR satellite systems, the

production of this type of classification map requires multiple images to be stitched together in order to create the large map.

This is a particularly difficult problem for VHR optical imagery because, in order to cover a large area, a significant number of images need to be included in a mosaic and a data model created that is portable between all images. While the precise requirements are problem specific, the differences between images include several noise and signal sources:

(1) Seasonality: land cover changes due to seasons

(2) Solar illumination angle: shadowing changes

(3) Atmospheric composition: scattering changes

(4) Collection geometry: observed atmospheric scattering due to the satellite look angle

(5) Observation angle: building lay-over

(6) Surface reflectance (BRDF): changes in observed reflectance due to view-angle

(7) Real surface changes: Natural and man-made surface changes

Due to these effects, LU/LC classification models have limited applicability outside of the image from which they were created. This is because the data space of the typical VHR image is distorted due to a number of physical and instrumentation effects. This means that a LU/LC classification model created for one VHR image would not be applicable to the additional images necessary to make a large area map. In order to address this limitation, it is necessary to devise a method to port the data model from the image for which it was created to the remaining images that make up the large image mosaic.

In the following experiments, the distortions present in VHR imagery are addressed using physical normalization techniques. These techniques lean on physical modeling and correction of solar illumination, instrumentation, and atmospheric effects. More detail on these techniques and the resulting data sets are provided in Section 5.2.1 below. The normalized data spaces

produced by the described procedures are then explored in order to determine methods to optimize the portability of a LU/LC model based purely on understanding the physical parameters and constraints associated with multispectral imagery.

While still complex, data space normalization reduces the degrees of freedom for a spectral land-cover classifier. This should, in theory, improve the generality of a classifier and improve its performance [91]. However, the data space normalization problem is particularly complex for VHR satellites, due to their capability to collect images at high-off nadir angles with associated solar illumination and observed atmospheric effects. This is further complicated by the temporal distribution (and accompanying seasonal and atmospheric changes) of composite multi-angle data.

In these model portability experiments, two types of multi-angle data are used to simplify the noise sources described above: in-track and composite multi-angle data. In each of these experiments, a multispectral machine learning model is trained using the information from a single image. This model is then ported across the multi-angle data space in order to directly and quantifiably measure the impact that data space normalization has on the model's portability. More details regarding the respective experiments are provided in Sections 5.2.1 and 5.3.1. The results of these experiments are presented in Sections 5.2.2 and 5.3.2.

5.2   In-Track Multi-Angle Model Portability

Machine learning model portability can be very difficult to investigate due to the multitude of noise and information sources that create confusion and shift the data space in various ways. However, in-track multi-angle data, such as that provided by the highly agile DigitalGlobe satellites WorldView-1 and WorldView-2, provides the unique ability to eliminate some of the complicating factors described previously.

This type of data set is a sequence of imagery (taken seconds apart) collected from multiple view-angles during a single overflight of the satellite. Therefore, there are no major changes in the composition of the atmosphere, solar illumination, seasonality, or on the surface between images. The only remaining effects that affect the portability of a multispectral LU/LC model are

short-term movement of vehicles (during the seconds between images in the sequence), changes in the atmospheric absorption and scattering due to view angle, and observed changes of surface reflectance due to BRDF. These simplifications of the full portability problem allow us to directly investigate physically based data space normalization methods (such as band ratios and atmospheric compensation) and quantify the ability of these methods to mitigate the view-angle and atmospheric effects.

In these experiments, the fundamental aspects of model portability are explored by measuring the ability to port a machine learning model, created at a single image of the in-track multi-angle sequence, to the other images in the in-track multi-angle sequence. By measuring the classification performance, a direct measure of the model's portability is created. In the presented experiments, this measurement is reported with Cohen's kappa coefficient for each image in the sequence and compared to the training location (where the model will provide maximum classification performance due to it's optimization for the raw data space of that image). In this way, the reported measurements are a literal and direct measure of the model's portability for the imagery and data space investigated.

In all cases, the LU/LC classification models use only the eight-bands of multispectral information collected by the WorldView-2 satellite. Additional information sources, such as surface texture, and their impact on portability are not treated in these experiments. The impact of data space normalization for the spectral machine-learning model portability are then measured by calculating the loss of classification accuracy across the image sequence for the same ground targets. Using this information, an analysis is presented of both the impact of data space normalization and collection geometry on spectral classification accuracy and model portability.

5.2.1   Methods: In-Track Multi-Angle Model Portability

The experiments presented in this section analyze this simplified data space of the in-track multi-angle sequence to test performance and characteristics of four data space normalization techniques. Each of these four data sets represents a different attempt (some easily calculated, some

more complex) to correct for error sources in the observation path. Together with the raw DN observations, this provides five data sets, all created from the Atlanta in-track multi-angle image sequence, that will be used in these experiments:

(1) Satellite observed digital number (DN)

(2) Top of atmosphere (TOA) reflectance

(3) Atmospheric compensation (AC)

(4) Band-ratios (BR) calculated from satellite observed digital number (DN)

(5) Band-ratios (BR) calculated from atmospherically compensated (AC) data

Digital Number (DN) is the satellite observation data space and is normally the data space in which a satellite image is delivered. In the case of the WorldView-2 imagery used in this experiment, the DN data space is composed of eight multispectral bands with 11-bit dynamic range. These DN values are the recorded energy intensity impinging upon the satellite sensor array. They are dependent on the instrumentation properties of the satellite as well as the integration time set during each collection. These values are also corrected for low level observational error (relative radiometric calibration) as is typical for higher level operational observations. However, for the practical exploration of the experiments, these DN values can be considered as the satellite observed energy intensity value. In these experiments, this will serve as the base case representing the typical "uncorrected" data space.

Top-of-atmosphere (TOA) reflectance data is corrected for the specific sensor characteristics, namely the spectral radiance response, instrument gain, band integration solar spectral irradiance, and solar irradiance geometry [21], to convert the DN data (pixel values between 0 and 2048) to a more physically meaningful reflectance (pixel value between 0 and 1) [1]. In this data space, a pixel with reflectance value of zero would represent a perfectly dark surface material and a pixel with a reflectance value of one would be observing a surface that is completely reflecting in the observed spectral band. This is the basic "physically meaningful" data space which, if perfectly executed,

provides a data value that would be the same for any sensor that observed the same scene at the same location.

Atmospheric compensation (AC) can take many forms and has been the subject of extensive research since the first Earth observing satellites were placed into orbit [5,62,70]. These techniques are designed to convert satellite digital number values to surface reflectance. Theoretically, this converts each observation into an instrumentation and observation angle independent data space that accurately represents the reflectance of the observed material as if the observation was made on the ground.

In this research, AC data are created using a rigorous inversion of the radiative transfer MODTRAN code. This system is the defacto standard of atmospheric modeling and has been implemented in several commercially available packages, such as FLAASH (*Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes*) [23], which is available in the ITT analysis package ENVI (*ENvironment for Visualizing Images*). Unfortunately these techniques are very computational complex and are just recently being developed for large scale deployment. Other less computational complex approaches have proposed and can be operationally useful, however, they often rely on strict imagery and target availability requirements to produce realistic output [14]. Therefore, the AC data used in these experiments is created using MODTRAN radiative transfer code because this presents the optimal combination of accuracy and operational prospects necessary for the full large-area model portability problem.

In addition, two band ratio (BR) data sets are created. These data sets are an attempt to compensate for the illumination and shadowing effects across the multi-angle sequence. For LU/LC classes with significant vertical structure, this can be a large effect [38]. While there are also material reflectance properties that drive the observed BRDF [53], in effect, the calculation of a band ratio is normalizing to the illumination and shadowing of a target area [74].

Both band ratio data sets are created with the same ratios, but used two different inputs: band ratios from digital number (BR-DN) and band ratios from atmospherically compensated data (BR-AC). In each case, the eight multispectral WorldView-2 bands are used to calculated a full set
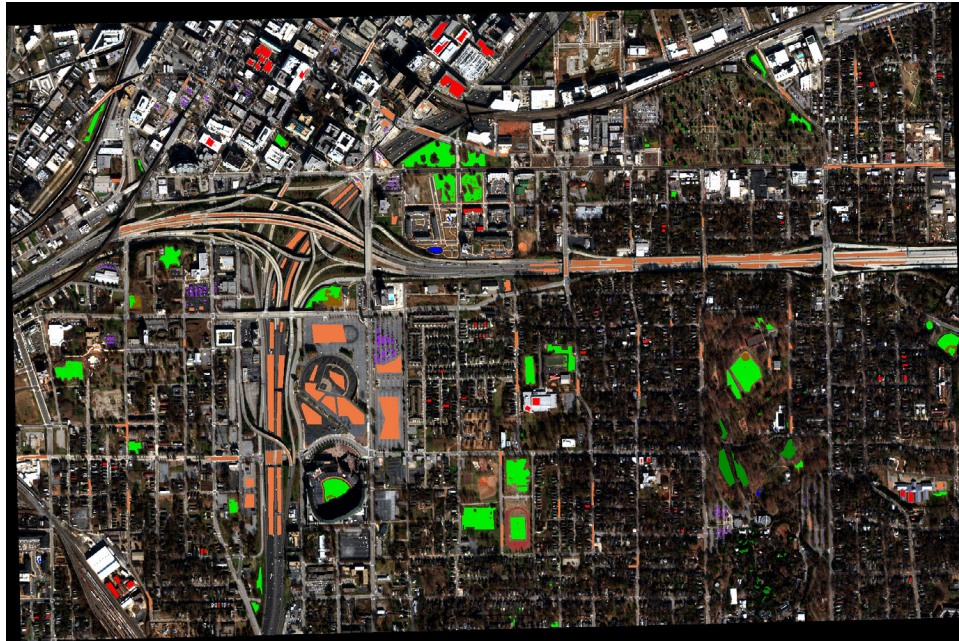
Figure 5.1: Ground survey used for the in-track multi-angle model portability experiments. The classes include: ■ = water, ■ = concrete, ■ = asphalt, ■ = vehicles, ■ = soil, ■ = grass, ■ = buildings, ■ = shadow, and ■ = trees. The same class selections were extracted from each of the 24 images of the composite multi-angle sequence.

of normalized NDVI-type (normalized difference vegetation index) ratios. These are calculated as $(band2 - band1)/(band2 + band1)$ with all iterations possible from eight band data. This includes ratios between bands 8–7 $((band8-band7)/(band8+band7))$, 8–6 $((band8-band6)/(band8+band6))$, 8–5, ..., 7–6, 7–5, ..., 3–2, 3–1, and 2–1 for a total of 28 ratio features in each data set.

The training and validation data for these experiments consist of nine land-cover classes: water, concrete, asphalt, vehicles, soil, grass, buildings, shadow, and trees. A total of ninety thousand pixels, illustrated in Figure 5.1, were selected through visual inspection for the ground survey. From this survey, 100 pixels per class (900 total) were extracted for training with the remaining pixels (approximately 89,000) reserved for validation. These same training and validation pixels were then extracted from each of images in the in-track multi-angle sequence, in each of the five data sets, for use in the portability experiments.

With the training and validation data extracted from each of the five data sets, thirteen model portability experiments are performed on each data set. These experiments are set up to directly

test the portability of a Random Forest machine learning model by porting a model, created from one image in the sequence, across the rest of the images in the sequence. This is then repeated for the remaining observations, creating a new model at each observation and porting it across the image sequence. In the case of the Atlanta in-track multi-angle data (see Chapter 3 for more details), this creates thirteen experiments executed for each data set:

(1) Training location: image 1; applied to all thirteen images.

(2) Training location: image 2; applied to all thirteen images.

...

(13) Training location: image 13; applied to all thirteen images.

An example of this direct portability test is shown in Figure 5.2. In this experiment, a Random Forest LU/LC model (represented by the blue/dashed box) was created using the training data from image observation #3 (red/solid box). This model was then applied to the validation data from observation #3 as well as the rest of the observations (see Figure 3.4 for observation and solar geometry information). The accuracy of the model, as applied to each image, is measured by Cohen's kappa coefficient or F-measure to create a direct measurement of the model's portability in the data space (in this case DN). Since the Random Forest model is non-deterministic [46], this experiment is repeated 30 times using the same training and validation data in order to gage the statistical significance of the results (see Appendix A). This creates a series of box plots that represent the statistical portability for a model trained at the geometry of the observation used to train the model.

These experiments are then repeated by training a model at each image in the sequence. After completing these experiments for all five data sets, results from a total of 65 experiments are available to describe the portability of a multispectral LU/LC model in a variety of data spaces. From each of these experiments, we generate a direct measure of the portability for not only the
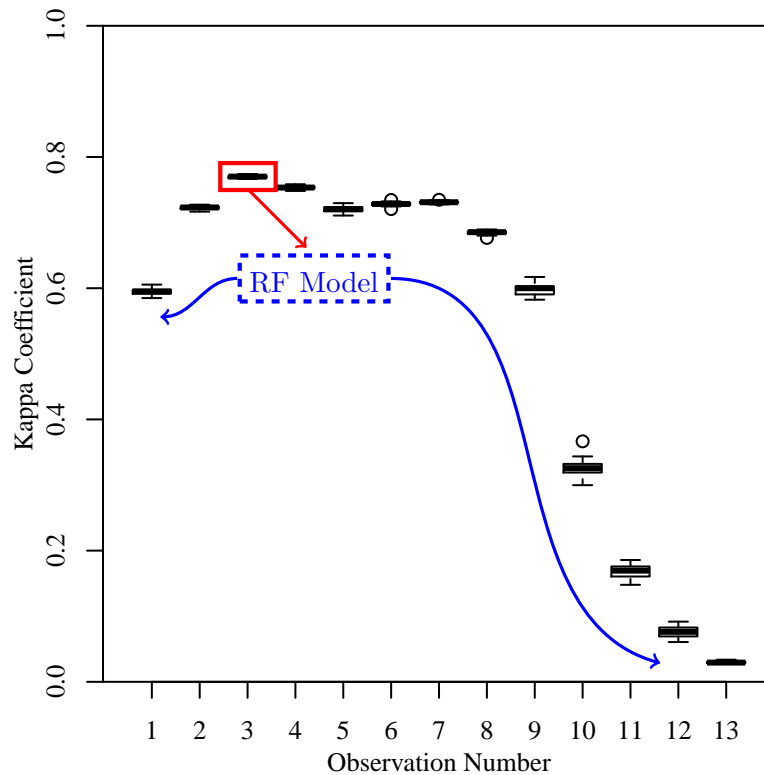
Figure 5.2: An illustration of the in-track model portability methodology. In these experiments, training data from a single image (red box) is used to create a Random Forest model (blue box). This model is then propagated to the validation data from the images in the multi-angle sequence. The classification accuracy (measured by kappa coefficient) is then reported for each image, and the experiment is repeated 30 times in order to build up a statistical certainty of the classification performance. This process is then repeated (not shown in the illustration), training a new model at each image location and testing the portability of that model across the multi-angle data.

overall model (measured using Cohen's kappa coefficient), but also for each class (measured using the F1-score).

5.2.2   Results: In-Track Multi-Angle Model Portability

For each of the five data sets explored in this study, three of the thirteen model portability experiments are presented in the plot panels of Figures 5.3 and 5.4. In the experiment shown in the left panels, a Random Forest model is trained at observation #3 (corresponding to a ground observed azimuth of 0°and elevation of 59°) and ported to the validation data of the remaining 12 observations. This experiment is then repeated with the model trained at observation #7
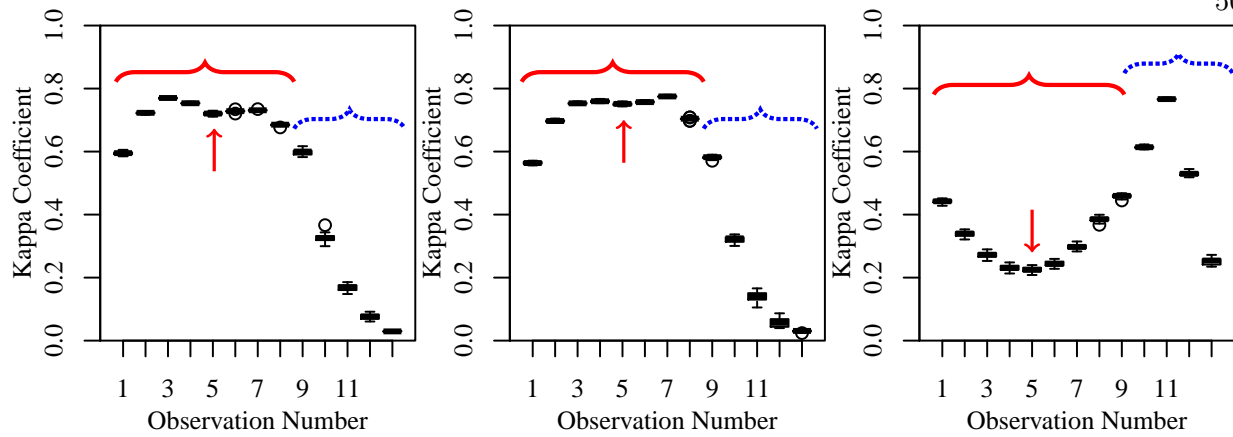
(a) In-track model portability in DN data space.
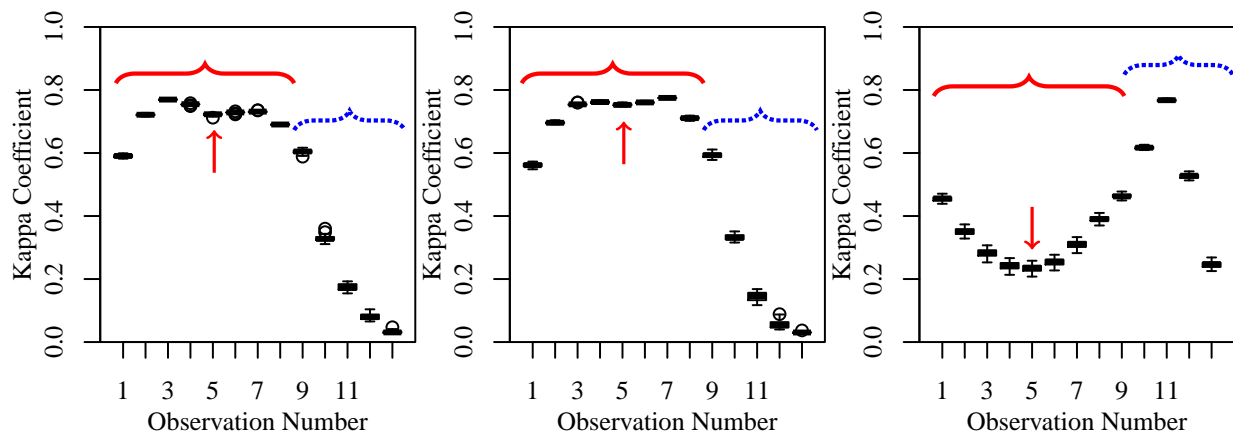
(b) In-track model portability in TOA data space.

Figure 5.3: Model portability experiments with, in each subfigure, the model train at observation #3, #7, and #10 and then ported to the other observations. A kappa coefficient is then calculated from the validation data sets. This experiment is then repeated 30 times to judge statistical significance of the model portability. These results are plotted in box plots for the raw digital number (DN) data space (a) and the top-of-atmosphere (TOA) data space (b). The red braces denote the solar forward scattering direction with the red arrow indicating the image in the sequence that is most directly opposite the target area from the sun. Conversely, the blue braces denote the backward solar scattering direction.

(ground observed azimuth=300°and elevation=82°) and #10 (ground observed azimuth=225°and elevation=72°) in the middle and right panels respectively.

While experiments were ran training a model at each image location, only three plots, with the model trained the image locations #3, #7, and #11, are presented because they are representative of the full range of results and provide a view of the portability behavior with the model trained
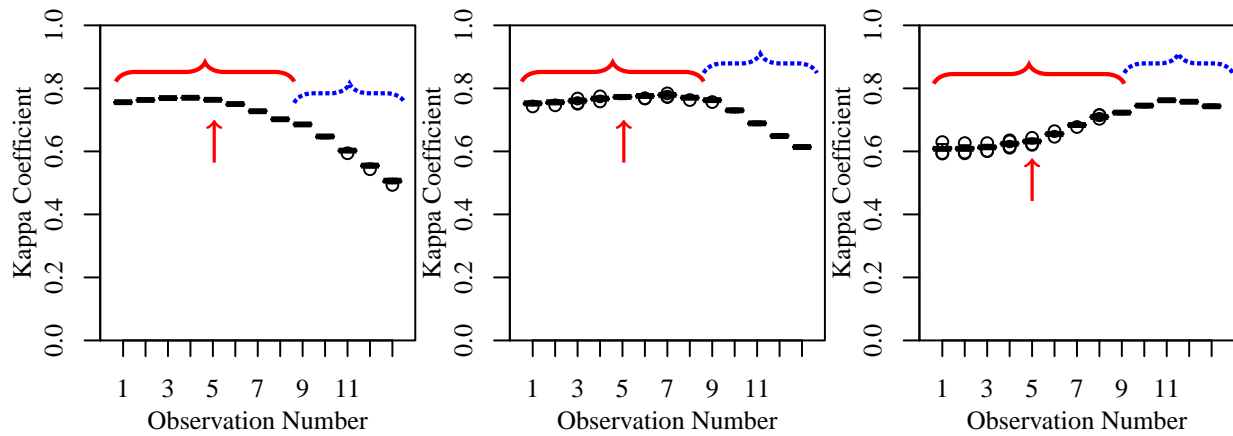
(a) In-track model portability in AC data space.
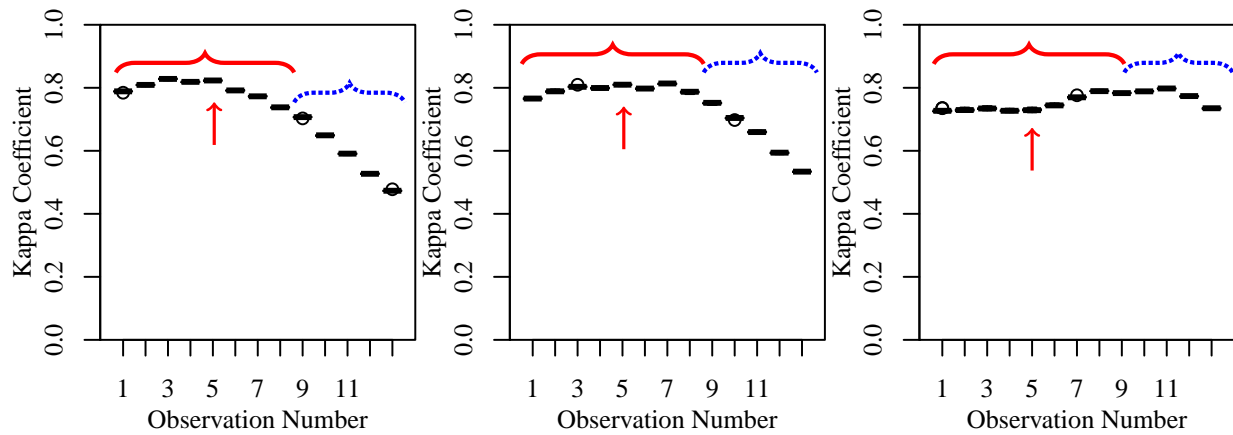
(b) In-track model portability in BR-DN data space.
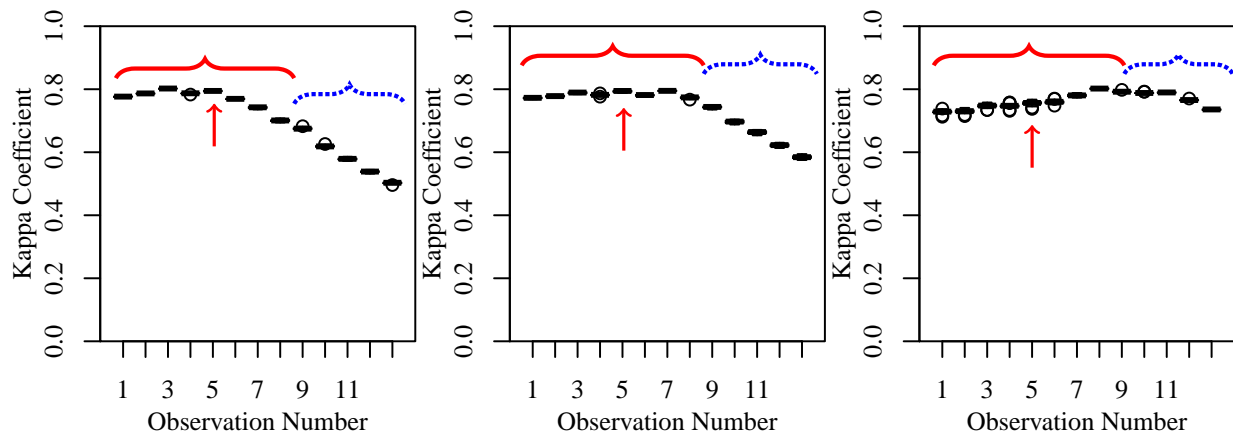
(c) In-track model portability in BR-AC data space.

Figure 5.4: The same model portability experiments, described in Figure 5.3, applied to the atmospherically corrected (AC) data space (a), DN band ratio (BR-DN) data space (b), and atmospherically corrected band ratio (BR-AC) data space (c).

at different locations along the image sequence. With a clear division between the forward and backward scattering regions in the DN data space, these can be best described in terms of the solar scattering angle. These angles are illustrated by the brace and arrow markers in Figures 5.3 and 5.4. Referencing Figure 3.4, image five is the nearest image to the direct forward scatter angle opposite the observation area from the sun and is denoted by a red arrow. More generally, images one through eight are in the forward solar scattering region, illustrated by the red/solid braces, with the remaining images nine through thirteen in the backscatter region and denoted by the blue/dashed braces.

We can see from Figure 5.3, that the multispectral model utilized in these experiments is providing a reasonable overall classification kappa coefficient. In each of the three panels, the Random Forest model was trained and validated at a single image location (#3 in the left panel, #7 in the center panel, and #10 in the right panel) and provides a classification result of of approximately 0.80 for each result despite the lack of additional derived features, such as texture or morphology, that are often used in this type of analysis. This is a reasonable performance level from which to compare the experiments of this chapter, providing a single image accuracy at or near levels considered state-of-the-art [22].

The DN data space provides insight into the effects that the dynamics of atmospheric scattering induced in non-corrected imagery. In this data space, there are large difference between the near-forward (obs. $\# \leq 8$) and high-backward (obs. $\# \geq 9$) scattering directions. For the observations within approximately $30°$(full angle) of the forward scattering direction, the ability to apply a model trained and applied within that cone is high. In the worst case, the kappa coefficient dropped to approximately 0.80. However, in the high-back scattering direction, there was no ability to port a machine learning model, regardless of training location. The same trends are also observed in the TOA dataset.

In both the DN and TOA experiments, there is a significant dependence on the training location in addition to the general trends just discussed. When a model is trained in the forward scattering region, there is reasonable portability across the same region. However, when a model

is trained in the backward scattering region, there is no model portability at all with the kappa coefficient falling off approximately 20% between the training image and the neighboring image in the sequence.

However, when applying data space normalization techniques to the model portability problem, the results change dramatically. As shown in Figure 5.4, the three normalization techniques show significantly improved model portability in all three experiments presented. This is evidenced by the elimination of the zero-kappa results at high-backscatter angles and the mitigation of the dip in kappa centered around image observation #5 (most pronounced in the right panel experiment). In all three of the presented data sets, the model portability is significantly improved compared to the raw DN and TOA data spaces with all kappa coefficient values above approximately 0.5 in the worst cases, and much higher on average.

As is the case with the DN and TOA experiments, the location where the model is trained is of significant importance. In the case where the model is trained in the forward scattering domain and applied across the full data sequence, the model portability improves, but still shows significant degradation in the high-backscattering domain with the lowest kappa coefficient of 0.60. However, in the normalized data spaces, the model portability favor training in the backward scattering region. When the model is trained in the high-backward scattering domain, the lowest observed kappa coefficient is approximately 0.65 across the 60°(full angle) observation cone of the multi-angle data.

In order to better compare the relative performance, the plots of Figures 5.3 and 5.4 are overlaid in Figure 5.6 and 5.5. To decrease the complexity of the plot, the markers delineating scattering angles have been left off the combined plots. Also note that the vertical scale of the kappa coefficients has been limited to a more illustrative range for the latter plot.

By directly comparing the DN (light red) and TOA (dark red) results (Figure 5.5), we can see that the two data sets provide statistically similar portability results. This is a reasonable and expected result given that the in-track multi-angle data set was collected during a single overflight of the satellite. During this time period it is highly unlikely that any of the pertinent
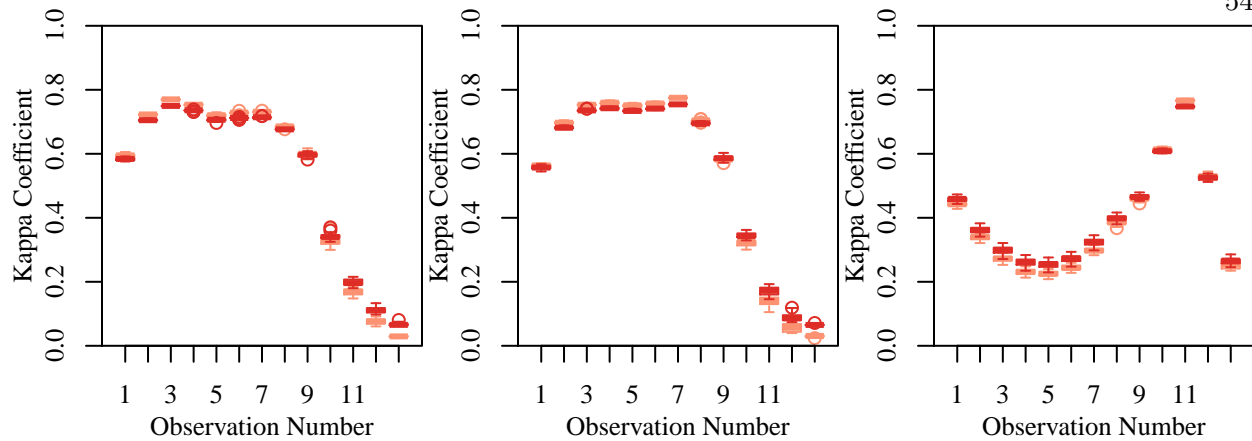
Figure 5.5: Overlaid model portability experiments of the DN (⬛) and TOA (⬛) data spaces.
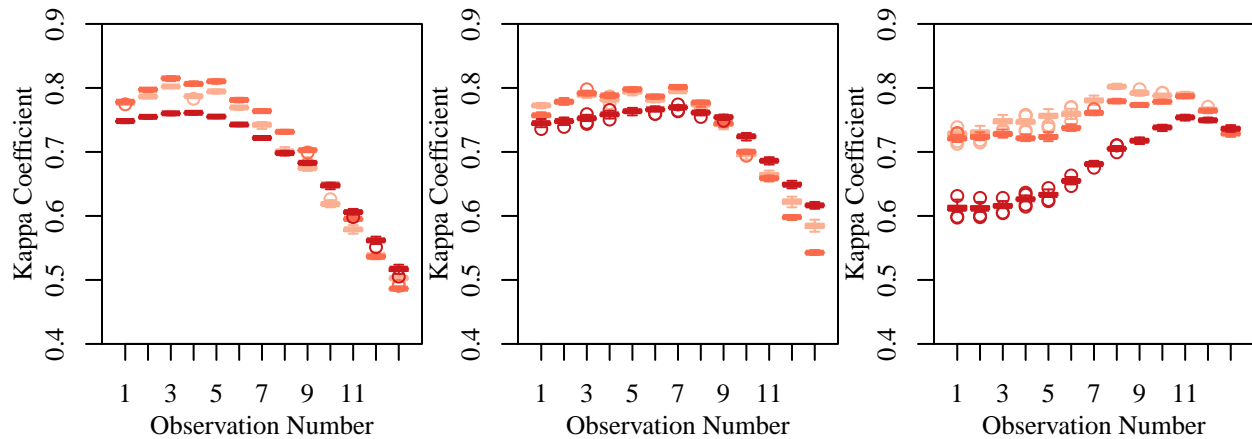
Figure 5.6: Overlaid model portability experiments of the AC (⬛), BR-DN (⬛), and BR-AC (⬛) data spaces.

satellite calibration setting or solar illumination conditions that are corrected as part of the TOA calculation would have changed. However, the TOA calibration could provide a useful normalization in multi-temporal data sets collected under varying atmospheric and illumination conditions.

Comparing the model portability results in the normalized data spaces presents more complex clues regarding the data space distortion effects that are addressed by the various normalization procedures. In Figure 5.6, the three data normalization techniques are shown in the same boxplots as Figures 5.3 and 5.4, but overlaid on an expanded kappa coefficient scale to show more detail. The three boxplot curves shown are the experiments using atmospherically corrected data (data

red), band ratios calculated from the DN data (medium red), and band ratios calculated from the atmospherically compensated data (light red). With the data plotted in this way, we can see small, but statistically significant, fluctuations between the three techniques that seek to create a more uniform, and therefore more portable, data space.

In these plots, three major trends can be identified:

(1) The AC data provides a smooth portability space with relatively elevated performance far away from the training location.

(2) The BR data sets outperform the AC data in the forward scattering region and near the training location, but are less smooth image to image and tail off more quickly away from the training location.

(3) When the model is trained the backscatter region, the portability across the full angle space is much more stable.

In this sequence of experiments, two different distortion effects are addressed: atmospheric scattering and illumination intensity. By combining data space corrections for these two effects(the BR-AC data space), the machine learning portability, with no active manipulation of the data space, goes from no portability (DN) to a maximum reduction of approximately (10%) across the full 60°test space. This is accomplished using only physically based corrections to address topographic, illumination, and shadowing as well as atmospheric scattering effects.

When appropriately performed, atmospheric compensation accurately removes the energy scattered at the sensor from the atmosphere as observed from the sensor's geometry. This is particularly important as the observation angle, and therefore the impact of atmospheric scattering, increases away from the training location. The impact of these corrections can be most clearly identified in middle and left panels of Figure 5.6. In these cases, despite a lower classification accuracy at the training location, the atmospherically compensated data space retains more portability far from the training location due to the removal of observation scattering distortion. This accounts for the largest improvements in portability observed as compared to the DN space.

However, atmospheric compensation is not designed to account for changes in illumination that are part of BRDF type effects as the observation angles changes across view-angle. As discussed above, previous studies have shown that band ratios provide both information enhancement and normalize to the illumination/shadowing conditions [111] of an image. Therefore, in essence, the band ratios correct for a subset of the full BRDF effects present in this image sequence. This effect can be observed in Figure 5.6 by the improved portability in the forward scattering regions. When the model is trained in the backward scattering region, where the objects are fully illuminated (and shadowing is minimized), the illumination correction maximizes portability across the observation angles of the in-track multi-angle data set.

These experiments show that neither the AC or BR normalization is completely sufficient to maximize the portability. The combination of these techniques provides a complimentary physical space normalization to maximize the portability. These conclusions are supported by the F1 plots of the individual classes. In the classes with significant vertical structure, such as trees (Figure 5.7) and grass (Figure 5.8), the band ratios provide a complimentary improvement to that of atmospheric compensation. However, for classes where there is little vertical structure, and therefore little variability of the types that band-ratios address, the atmospheric compensation maximizes the portability absent the creation of band-ratios. This is the case for the asphalt (Figure 5.9) and concrete (Figure 5.10) classes.

The experiments shown here suggest that good image-to-image model portability can be achieved through data space normalization and careful selection of the observation geometry used to train the model. In the presented experiments, a classification kappa coefficient reduction of less the 5% was attained across the majority of the 60°(full angle) observation cone by combining the band ratios with atmospherically compensated data and training at the largest back-scattering angle (right panel of Figure 5.3).

If we assume that the normalized data spaces shown here have accurately compensated for the atmospheric and illumination effects, then the remaining effects will be due to real BRDF changes. The size of this effect is greatly dependent on the specific class under consideration,
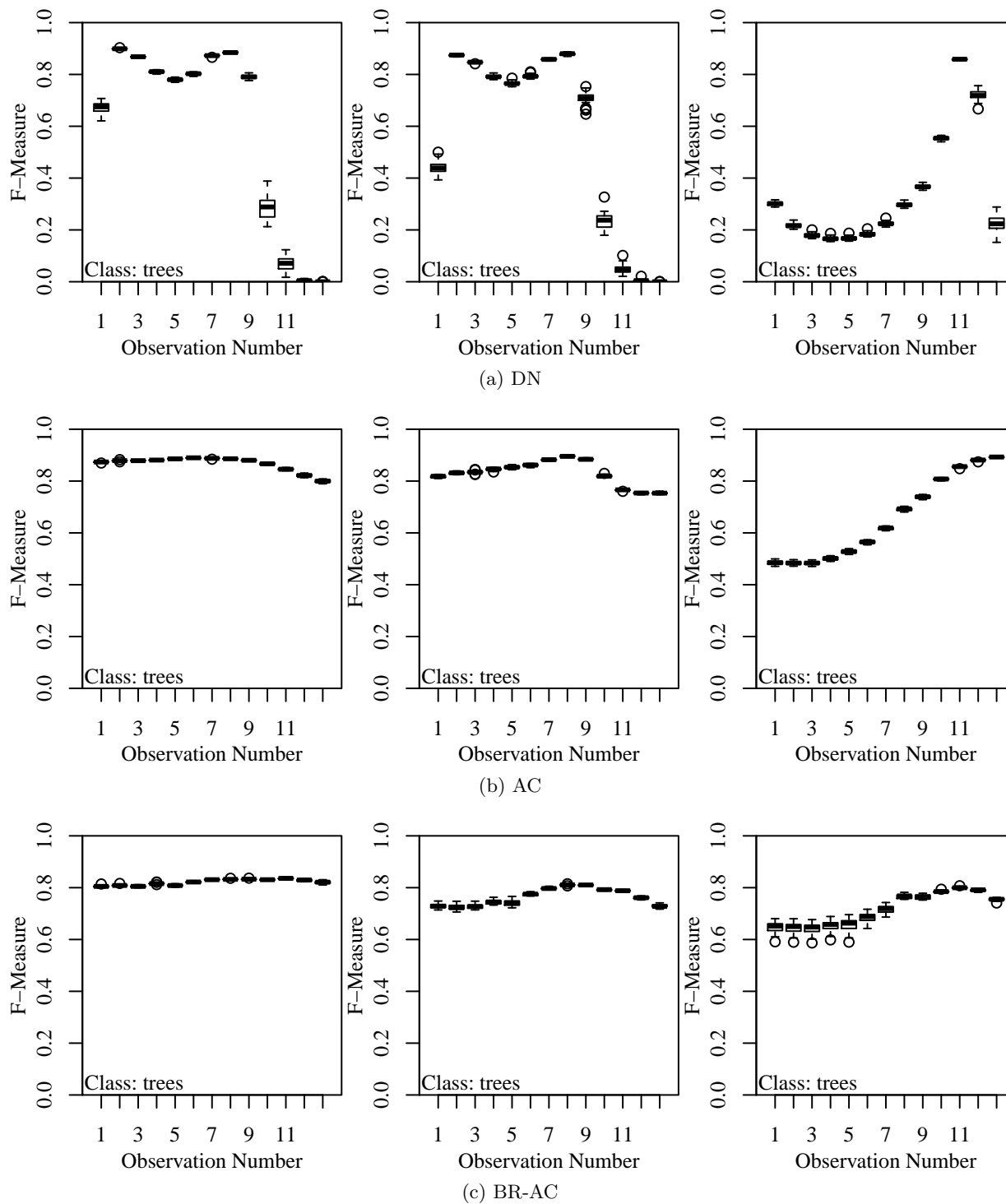
Figure 5.7: Model portability for the class "trees" in the DN, AC, and BR-AC data spaces.
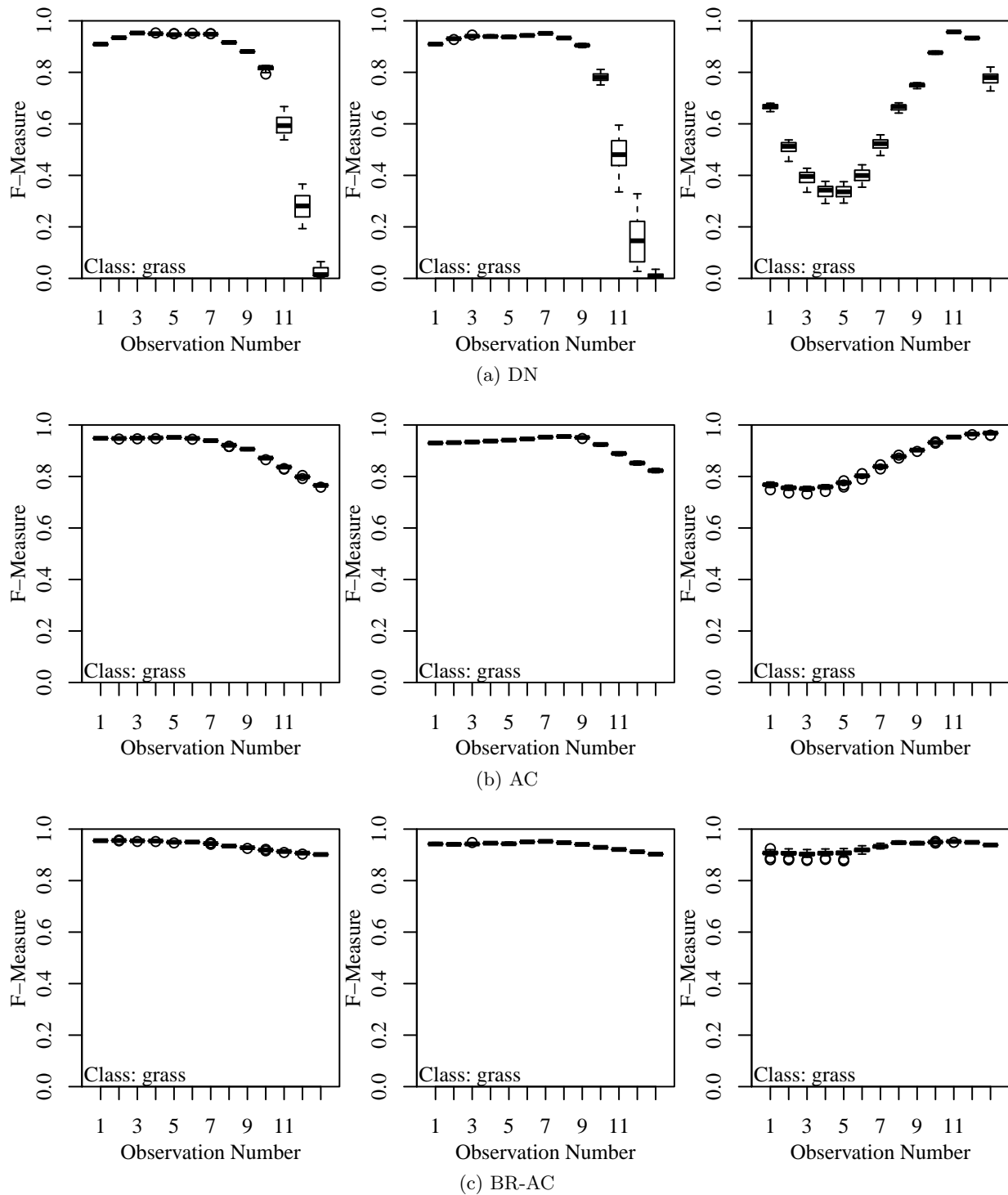
Figure 5.8: Model portability for the class "grass" in the DN, AC, and BR-AC data spaces.
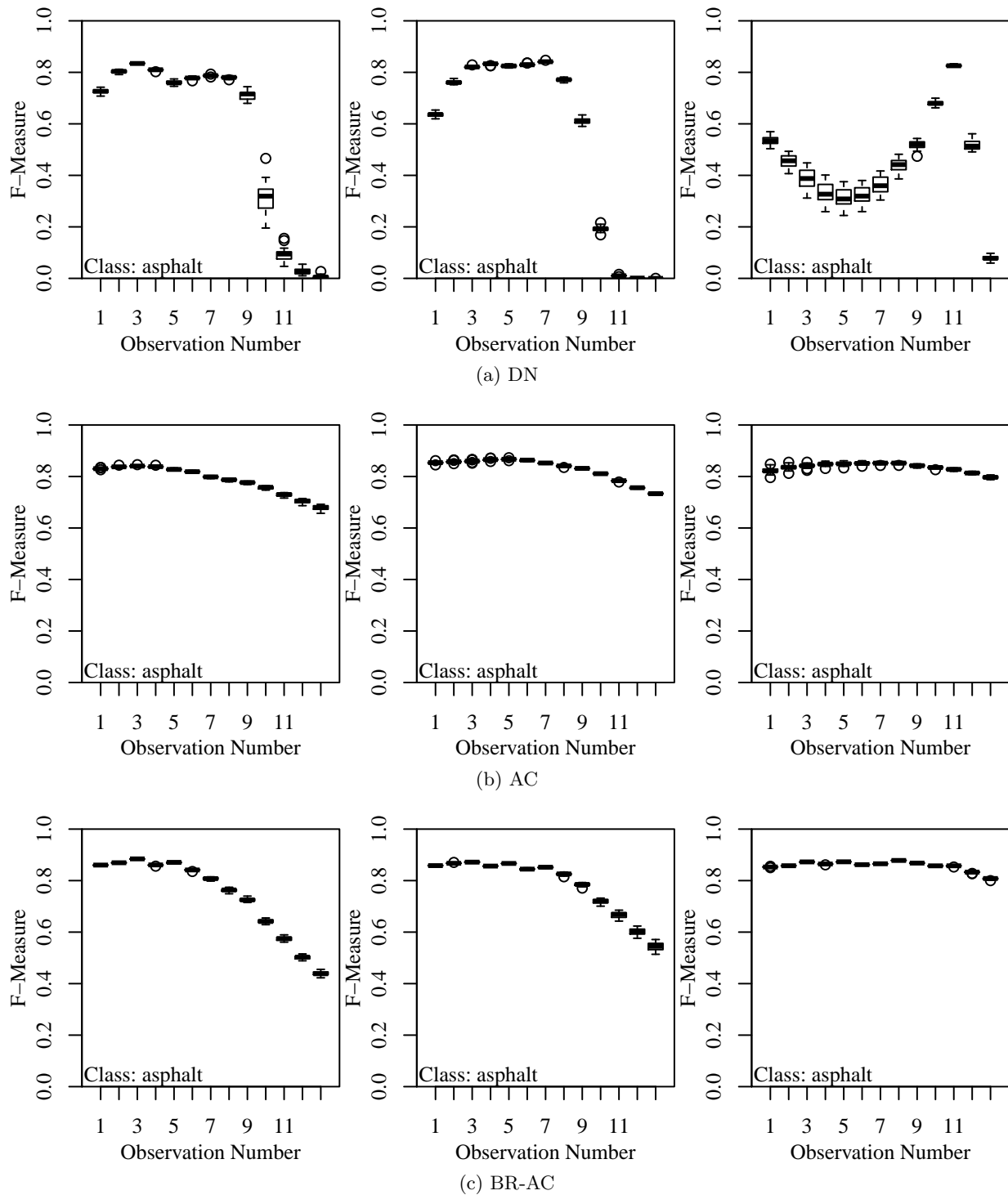
Figure 5.9: Model portability for the class "asphalt" in the DN, AC, and BR-AC data spaces.
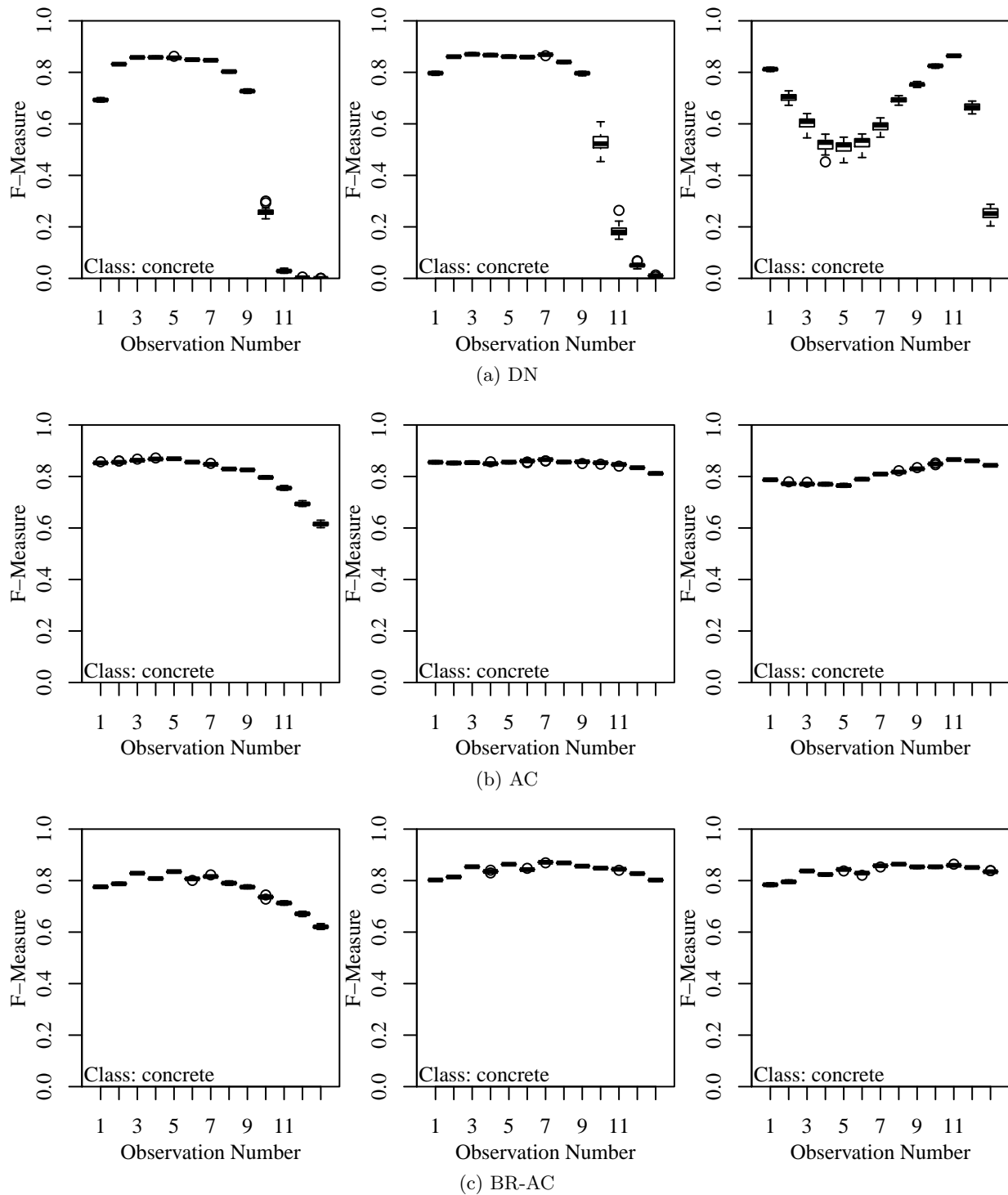
Figure 5.10: Model portability for the class "concrete" in the DN, AC, and BR-AC data spaces.

but, for the overall classification metric of kappa coefficient, the correction of these physical effects provides a data space where the effects of BRDF and seasonality can more completely be addressed. This creates a problem space that can form the foundation of solutions to the broader large area multi-temporal model portability problem.

Therefore, the next steps in this line of research would be to investigate methods of addressing these BRDF signals. This could be accomplished through either direct physical modeling of the BRDF of each surface [93] or through active learning techniques that seek to drive the spectral distribution in multiple images together by making assumptions regarding the overall structure of the data space [50, 84]. The effectiveness and approach could vary greatly by class. In the case of vehicles, there is no direct method to model the BRDF since there is no standard spectral signature of a vehicle. For the same reasons, active learning techniques could also fail. This may require the creation of contextual information (texture, morphology, etc.) for the accurate and portable classification of vehicles.

5.3   Composite Multi-Angle Model Portability

As informative as simplified test cases can be, it is necessary to move the in-track results toward the realistic parameters of a real model portability problem. The next step toward this goal is presented in this section using composite multi-angle data. The lessons learned from the in-track model portability experiments are extended to the more complex composite multi-angle data set.

In the composite multi-angle data set (described in Chapter 3), several of the simplifications present in the in-track case are no longer present. Since the images are collected over a period of nearly two years, the simplifications of seasonality, solar illumination angle, atmospheric compositions, and real surface changes are removed. This provides the opportunity to study a more realistic model portability case while preserving the ability to measure the impact of moving a model between images by observing the effect at the same surface location in each of the composite multi-angle images. In this way, the fundamental limitations and leverage points of the model

portability problem described by the in-track experiments are explored in a more complex data space representative of a large-area model portability problem.

This investigation will explore the ability to appropriately compensate for the data space distortions present in the composite multi-angle sequence through the application of physical data space manipulation techniques. This provides an opportunity to explore the lessons learned from the in-track experiments to the more generalized case of the composite multi-angle data. In this section, similar methods to those employed on the in-track multi-angle data are applied to the composite multi-angle data.

5.3.1   Methods: Composite Multi-Angle Model Portability

The experiments conducted in this section mirror those of the in-track model portability experiments. They explicitly test the portability of five normalized data spaces by applying a single machine learning model across the full composite multi-angle sequence. As in the previous experiments, the five data sets explored here are:

(1) Satellite observed digital number (DN)

(2) Top of atmosphere (TOA) reflectance

(3) Atmospheric compensation (AC)

(4) Band-ratios (BR) calculated from satellite observed digital number (DN)

(5) Band-ratios (BR) calculated from atmospherically compensated (AC) data

The development of these data sets mirrors that of the earlier model portability experiments (see Section 5.2.1) except that these data sets are created from the composite multi-angle data set. This data set was described in detail in Section 3.2.

From these data sets a ground survey of 18,000 pixels was selected including the following classes: asphalt, buildings, water, trees/shrubs, and grass. These classes are relatively static, so the experiments do not attempt to address dynamic changes on the surface through the time sequence.

Figure 5.11: Ground survey used for the composite multi-angle model portability experiments. The classes include: asphalt ▮, buildings ▮, water ▮, trees/shrubs ▮, and grass ▮. The same class selections were extracted from each of the 24 images of the composite multi-angle sequence.

These classes also have significant BRDF properties that are representative of the basic classes required for a large-area model portability project. The survey regions are shown in Figure 5.11. From this survey, 100 pixels per class are selected from each image for training, with the remaining pixels reserved for validation.

As with the previous experiments, a Random Forest machine learning model is created from the training data of a single image and is ported to the remaining images in the data set. The

model is retrained and ported a total of 30 times in order to gauge the statistical significance of the resulting portability measurements. This portability experiment is then conducted for each image in the data set resulting in a total of 24 portability experiments for each of the data sets (DN, TOA, etc.).

The analysis and visualization of the composite multi-angle model portability experiments are presented significantly differently. This is necessary because the multi-angle data consists of images spaced throughout the observational hemisphere. The resulting model portability classification performance can not directly be displayed on a two dimensional plot. Therefore, a more complex three-dimensional surface representation is created for each portability experiment.

In order to visualize the model portability space of the composite multi-angle date, a portability surface is created for each experiment. In each surface plot, an average of the 30 statistical samples of each location's classification performance, determined by kappa coefficient, is created. The image location of the Random Forest model training set is marked with a box. An example of this is shown in Figure 5.12a. From this distribution, a thin plate spline [44] is fit to create a surface representing the portability across the full multi-angle sequence. This is shown in Figure 5.12b. Finally, in order to better visualize the variations across the surface, the fitted surface is extrapolated throughout the observational hemisphere (shown in Figure 5.12c). This extrapolation comes with the standard caveat - it is subject to data trends within the sample space as well as the choice of fitting parameters and does not necessarily represent absolute portability at an arbitrary location in the observational hemisphere.

These portability plots are created (although not shown) for each of the 24 images in each of the five data sets. A total of 120 portability surfaces are created that, despite the usefulness for visualization, do not directly provide insight into the effects that drive the model portability.

Therefore, in addition to the surface plots, the portability of each classification model is also treated statistically as a distribution of kappa coefficients. For example, in the experiment described in Figure 5.12, the kappa coefficient of each of the classification experiments varies from a minimum of 0.41 (near the forward scattering direction) to a maximum of 0.90 (where the model
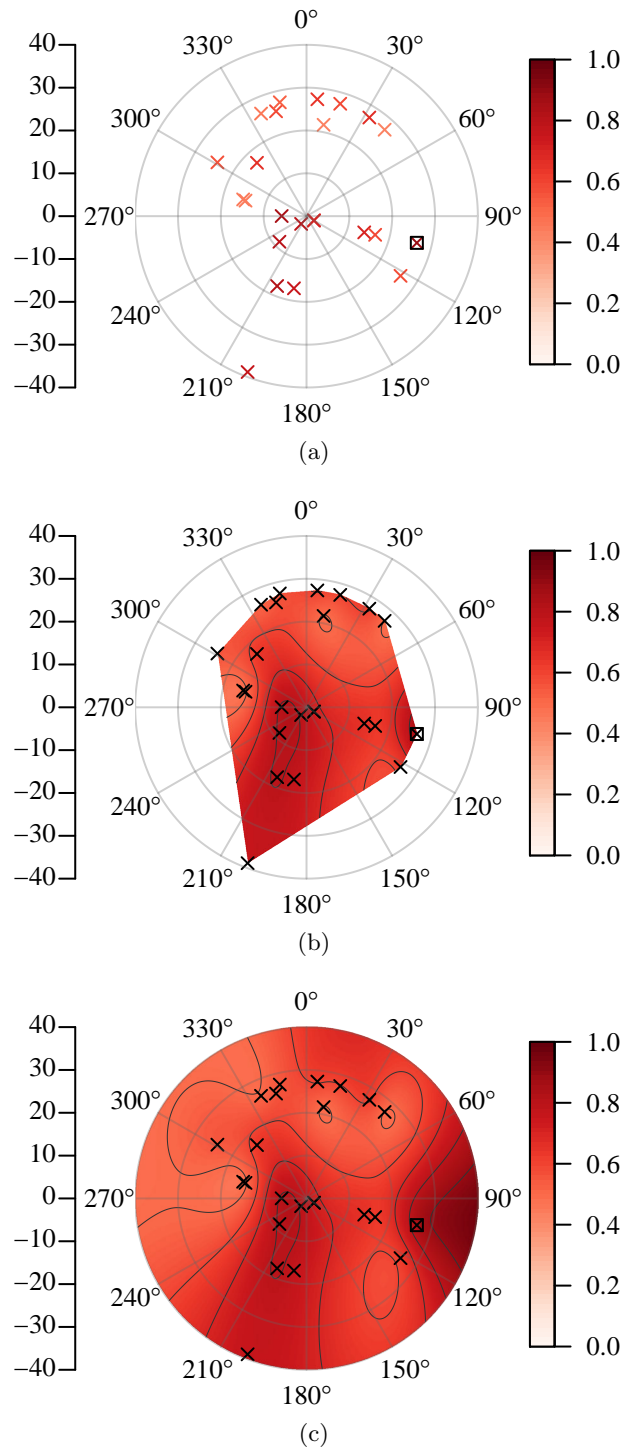
Figure 5.12: Individual components of the portability surface plots.
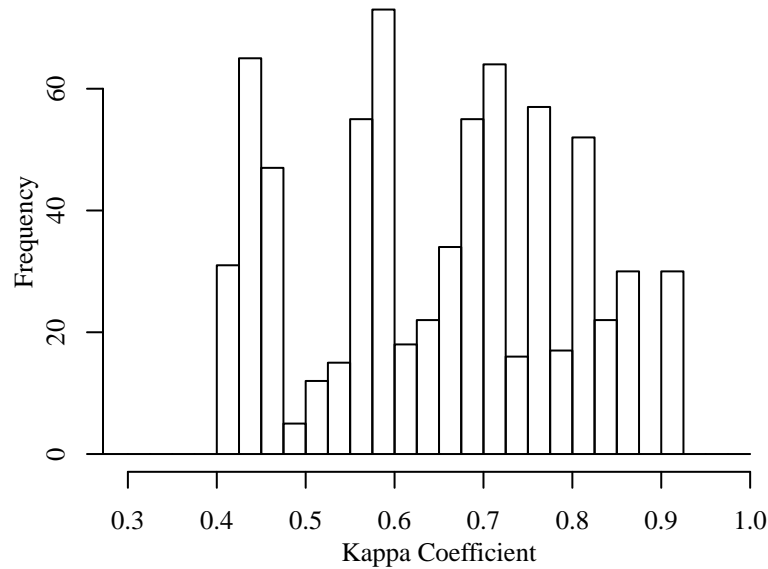
Figure 5.13: A histogram illustration of the kappa coefficient distribution (0.025 bin width) for all 720 composite model portability experiments for a single model training location (one model trained at a single location and ported to all 24 images, repeated 30 times to test statistical significance of the measurement).

	Mean	Median	Std. Dev.	Mode	Skew	Kurtosis
Example MP Exp.	0.65	0.66	0.14	0.59	-0.049	-1.1

Table 5.1: An example of the statistical metrics calculated for the model portability tests.

was trained). The histogram of the full range of kappa values that make up the portability surface is shown in Figure 5.13. From this distribution, portability can be described statistically by mean, median, mode, standard deviation, skewness, and kurtosis. These measurements, calculated for the example model portability exercise, are shown in Table 5.1. These values are then calculated for each exercise (a model trained at every image location) and for all data spaces.

Using these statical metrics, the overall distribution of each portability exercise can be illustrated. Three types of plots using these metrics are shown in Figures 5.14, 5.15, and 5.16. In the first two figures, the mean of each kappa distribution (as illustrated in Figure 5.13 and Table 5.1) is shown plotted by time of year and solar scatter angle. The errors bars denote the standard deviation of the distribution for each exercise where a model is trained at the given location (time of year or scatter angle) and ported to the remaining observation. In the relevant plots, the so-

lar scatter angle is defined as the central angle between the solar and satellite positions (azimuth and elevation) at the observation used to train the classification model. In Figure 5.16, various statistical metrics are provided in a scatter plot by data space types (DN, TOA, AC, etc.).

5.3.2   Results: Composite Multi-Angle Model Portability

The model portability results from the in-track data show that there are three effects driving the portability of a model:

(1) Atmospheric distortion

(2) Illumination changes

(3) Model training data space

The in-track results show that these effects can, to a large degree, be addressed through the use of physically based data space compensation mechanisms.

In additional to these effects, the composite multi-angle data has a significant seasonality component. The effects of this on model portability can be seen in Figures 5.14a and 5.14c where the model are more easily ported to other images from the same season. This is seen in the mean kappa value because the monthly distribution of the data sequence is weighted toward the summer season (see Figure 3.9). Therefore, since the models trained in the summer season are more easily ported to other summer images, the overall kappa mean is higher for the summer data points (and conversely lower for winter images).

In fact, this effect remains a complicating factor throughout the presented data sets. As shown in Figure 5.15, the same seasonality effect drives model portability with the same seasonal kappa mean correlation despite illumination and atmospheric compensation.

We can see, however, that the data space normalization has a significant effect on the model portability in the composite multi-angle data space. The normalization effects that can be measured in the following experiments include portability improvement due to sensor characteristics correc-

(a) DN: Kappa by Month of Year

(b) DN: Kappa by Solar Scattering Angle

(c) TOA: Kappa by Month of Year

(d) TOA: Kappa by Solar Scattering Angle

Figure 5.14: Time series and solar scatter angle plots for each DN and TOA portability exercise. Each measurement is the mean kappa coefficient of a model ported to all composite multi-angle images when a classification model is trained at the given location. The error bars represent the standard deviation of the kappa distribution of each location.

tion (TOA data space), atmospheric compensation, illumination correction, and model training orientation.

In Figure 5.14, the top two panels show the average kappa distribution (with one sigma standard deviation) of the portability experiments performed with a model trained in the DN and TOA data spaces at the marked location (date in Figure 5.14a and solar scattering angle in Figure 5.14b). The difference between the two sets of plots is small, but amounts to an improved

average portability (as measured by mean kappa) of 0.07 kappa or 11% improvement over the DN data space.

This result is opposite that of the in-track model portability experiments. As discussed previously, the TOA normalization provided no significant portability improvement when the images were collected during a single over-pass of the satellite. However, as is the case in the composite multi-angle data, the shift from DN to TOA compensates for variation in the satellite observed data space. In this present case of model portability, this produces a significant normalization from the completely uncorrected DN data and a correlated improvement in classification model portability.

The first normalization level explored in the in-track experiments, atmospheric compensation, is shown in the same kappa mean plots for the composite multi-angle data in Figure 5.15a and 5.15b. These plots show a significant portability improvement over the TOA data space. This is due to the compensation of atmospheric distortion throughout the data set, which takes the form of scattering differences due to observation angle as well as changes in the makeup of the atmosphere between collection dates. This is evidenced by both the 0.04 improvement in mean kappa as well as the reduction of 0.02 in the average stand deviation across the 24 experiments.

Portability results for correction of the illumination variation using band ratios (as described in Section 5.2.1) are shown in Figures 5.15c and 5.15d for the DN data. The two BR-DN plots show a subtle improvement from the portability of the raw DN space. In the BR-DN experiments, the outliers from the DN experiments have been pulled closer to the average distribution. This is evidenced by a reduction of the variability of the kappa mean values across the 24 exercises from 0.09 kappa to 0.05 kappa.

However, as with the in-track data, maximizing the portability is a function of normalizing for both the atmospheric and illumination effects. Figures 5.15e and 5.15f illustrate that the BR-AC data space both maximizes the mean kappa classification accuracy as well as minimizes the kappa variability of the 24 portability exercises. It is with these normalizations that the seasonality effects are minimized, improving the overall portability kappa mean to 0.77 ± 0.07 (standard deviation of kappa distribution).
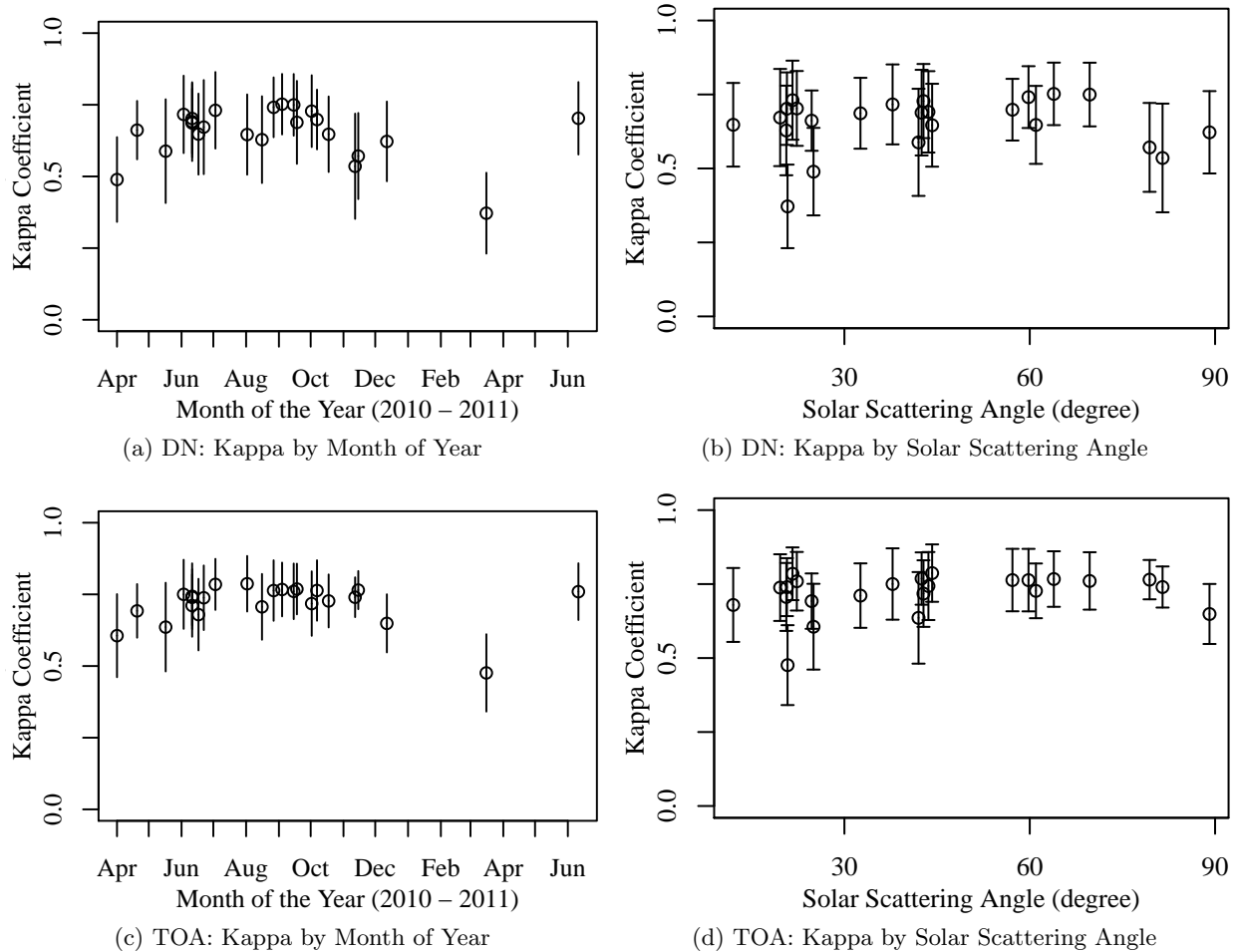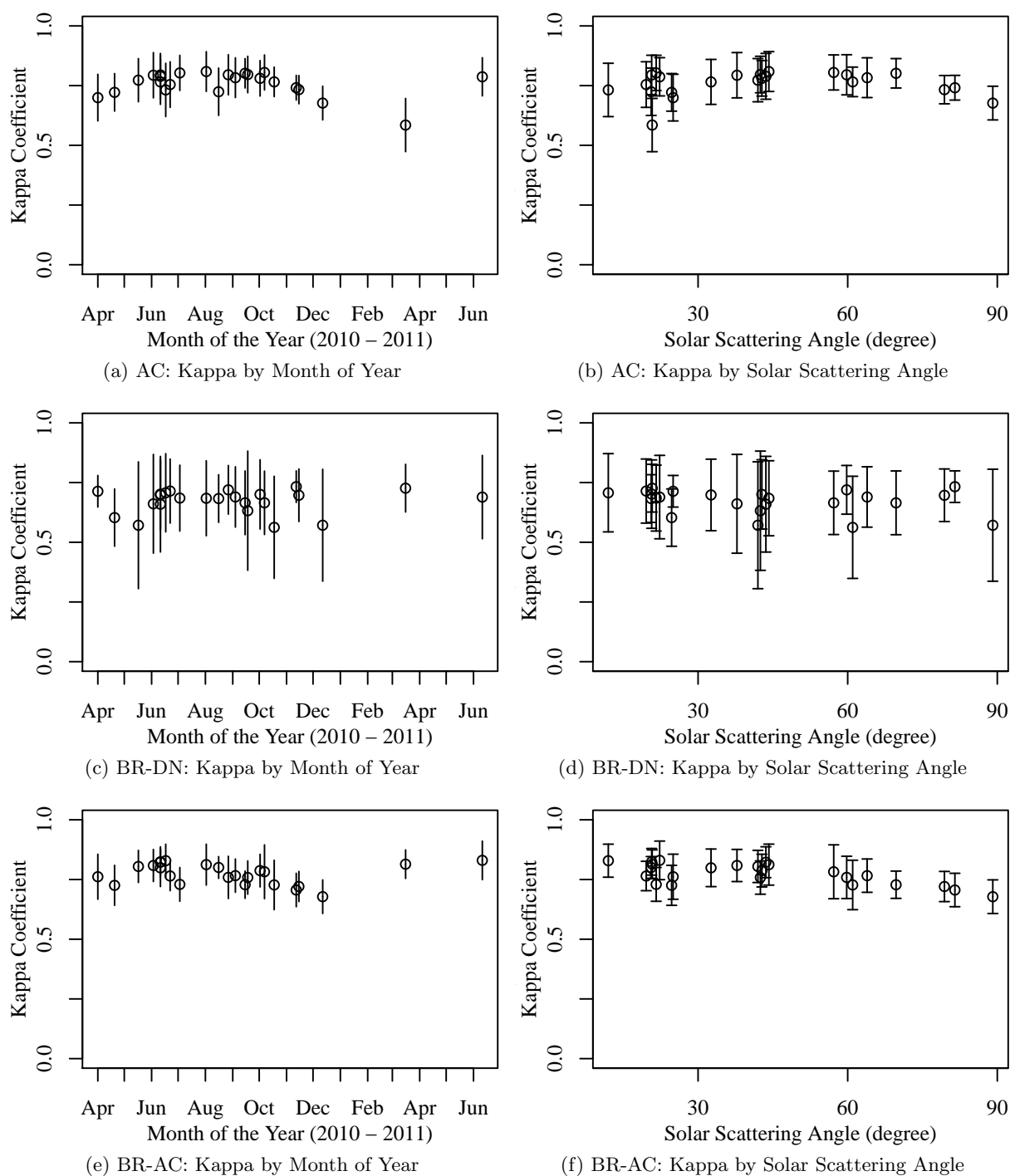
Figure 5.15: Time series and solar scatter angle plots for each AC, BR-DN, and BR-AC portability exercise. Each measurement is the mean kappa coefficient of a model ported to all composite multi-angle images when a classification model is trained at the given location. The error bars represent the standard deviation of the kappa distribution of each location.

Finally, as discussed in the in-track experiments, the location where the model is trained can play a role in the portability performance. This effect is not present in most of the composite multi-angle data. However, in Figure 5.15f, a small portability dependence on solar scattering angle can be observed. Fitting a linear model to this data (fit parameters are: slope=-0.0013; y-intercept=0.83) shows a statistically significant reduction in mean kappa accuracy across the solar scatter angle range of 0.1 kappa.

In addition, three scatter plots of the key statistical metrics for the portability experiments are shown in Figure 5.16. In each of these figures, the mean of the kappa distribution (the y-axis of the previous plots) is shown plotted against the standard deviation, skewness, and kurtosis of the kappa distributions. In each case, the most favorable statistics are:

(1) High kappa mean/low standard deviation

(2) Small or negative skew

(3) Large kurtosis

The reasoning behind the desired higher kappa mean is obvious - this bears a direct relationship to better classification performance. Similarly, a lower standard deviation is related to a more compact kappa distribution and, therefore, a more portable model. Skewness is a measure of the symmetry of the distribution. In the portability experiments discussed here, better portability performance is related to skewness by either no skew or skewness toward the higher kappa (lower skewness values). Conversely, higher kurtosis (a measure of the distribution peakedness) will correlate to a more consistent kappa coefficient as a model is ported across the images, and, therefore, a more portable model.

With these parameters in mind, the three scatter plots point toward similar conclusions to those drawn from the date and solar scatter angle plots:

(1) Atmospherically compensation provides a necessary normalization against the multi-angle and multi-temporal atmospheric effects.
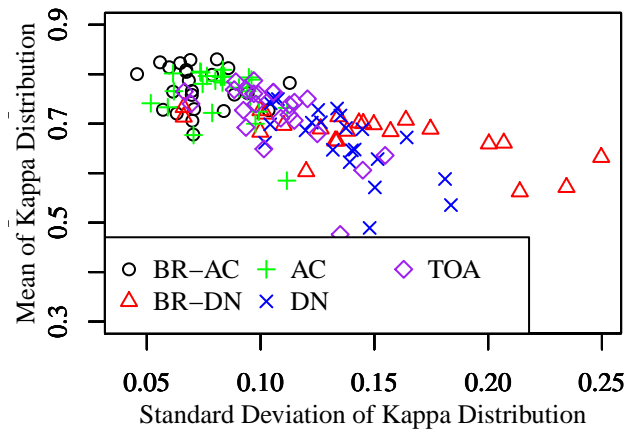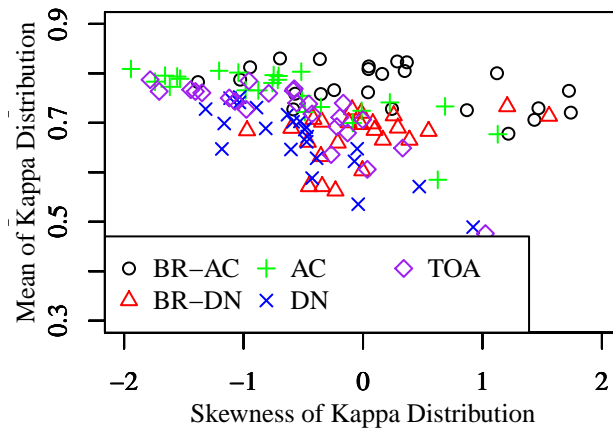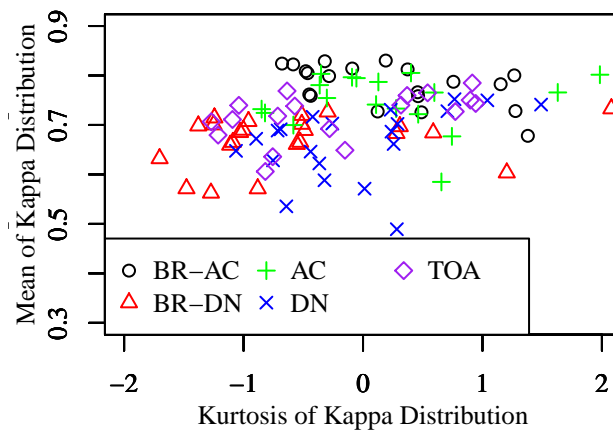
Figure 5.16: Scatter plots of the composite multi-angle model portability statistical metrics (mean, standard deviation, skewness, and kurtosis) for all DN, TOA, AC, BR-DN, and BR-AC portability exercises.

(a) Kap=0.82; SD=0.056; Skew=0.29; Kurt=-0.68

(b) Kap=0.83; SD=0.07; Skew=-0.36; Kurt=-0.32

(c) Kap=0.80; SD=0.046; Skew=1.1; Kurt=1.3

(d) Kap=0.81; SD=0.060; Skew=0.049; Kurt=-0.090
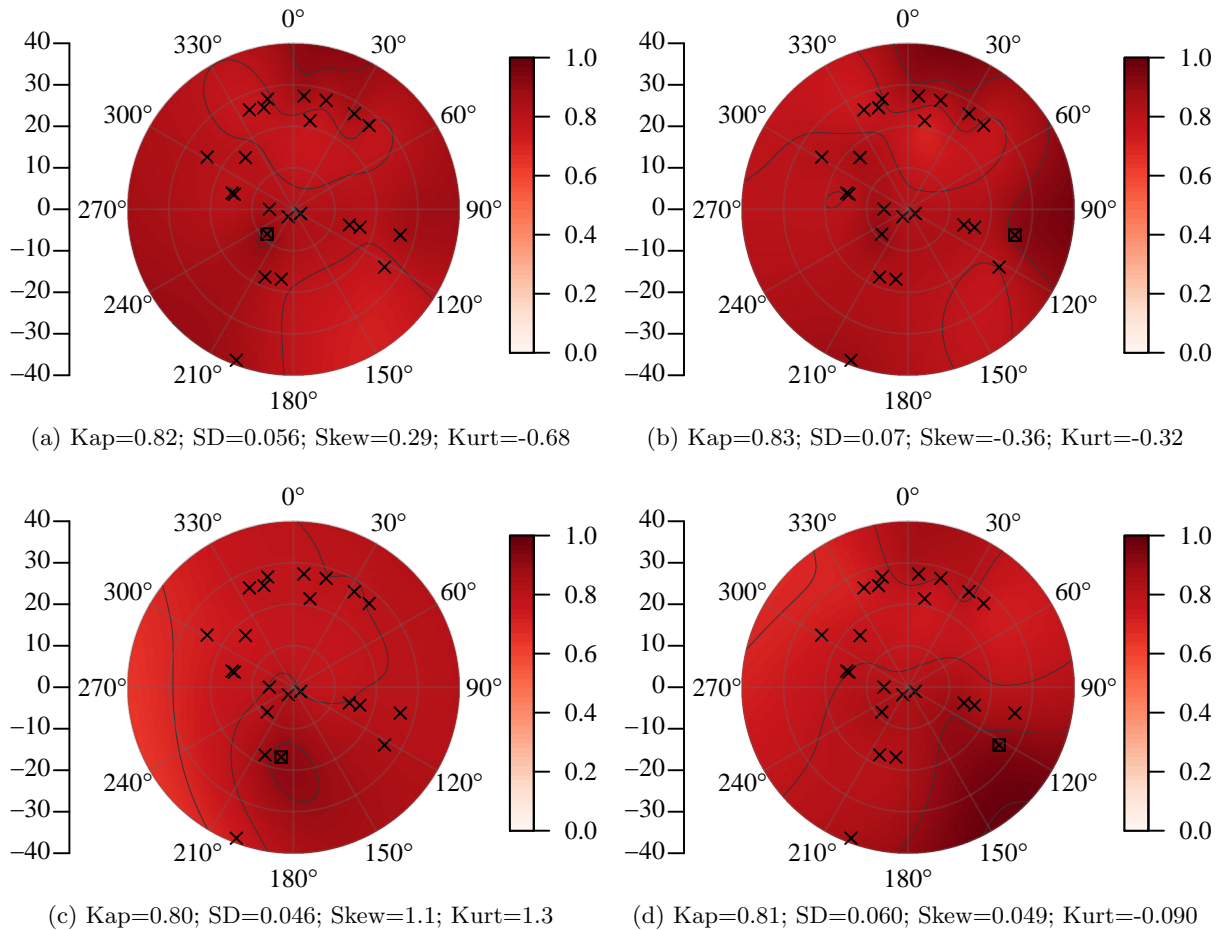
Figure 5.17: Selected example portability surfaces from the BR-AC data set. In time order (as plotted in Figure 5.15e), the models used to create the portability surfaces were trained at observation #6 (a), #8 (b), #12 (c), and #23 (d).

(2) Illumination compensation additionally improves model portability as seen in higher kappa mean and lower standard deviation (lower kurtosis).

When these conclusions are considered, as well as the observed solar scattering angle preference, the portability surface improves significantly from that observed in Figure 5.12c. Four examples of these portability surface plots are shown in Figure 5.17. All of the subplots are from the BR-AC data set shown to provide the most favorable normalization conditions. In each of them, the portability is shown by the flat (portable) and dark (higher kappa) surfaces. As before, the location used to create the model is delineated by a box around the image marker.

The composite multi-angle, compared to that of the in-track multi-angle data, is significantly more complex due to seasonality, illumination/observation geometry, and changes in the atmospheric composition over time. Despite this, the portability results generally follow the conclusions drawn from the in-track experiments in that the same effects driving classification model portability can be addressed by sequentially applying atmospheric compensation and band ratio illumination correction.

In addition, a small preference for training in the backscatter direction is observed. Theoretically, this effect can be explain be the fact that the "true" data space is more intact in the backscattering direction where the sun more directly illuminates the surfaces of interest. This, combined with the atmospheric compensation and illumination normalization of the band ratio calculations, creates a physically optimized portability data space. This data space increases the model portability of these experiments from no portability in the raw DN space to a scheme in which the loss in classification accuracy across a multi-temporal multi-angle data is as small as 0.21 kappa (maximum to minimum) with a standard deviation around the mean of 0.05 kappa.

In addition, the most basic data space normalization technique, top-of-atmosphere reflectance, has been shown to significantly reduce the distortions present in raw satellite data. This result is of general importance to the remote sensing community. In most experiments published in academic literature using optical satellite data, the authors utilize the DN space of the satellite data as the operational data space. While this is a reasonable approach when only a single image (or multiple images taken over a very short period of time) is used, a simple conversion to TOA reflectance is a minimal requirement in order to appropriately generalize any experimental methods involving satellite data.

Despite the physical data space corrections shown in this research, a small temporal signal can be construed in the final temporal portability plot (Figure 5.15e). The driving force behind this variation is most likely seasonality causing real spectral shifts in the classification data space (i.e. grass in summer vs winter). It is possible that this type of effect could be mitigated by moving

outside the physical normalization techniques that have been explored here. Such approaches could include active modeling or data space matching.

Chapter 6

Multi-Angle LU/LC Classification

This chapter details the experimental methods and results of the multi-angle LU/LC classification experiments. These experiments explore the unique aspects of WorldView-2 multi-angle data by combining very high spatial and multispectral resolutions with information extracted from multi-angle observations for the purposes of increased classification performance. The value of this information for urban classification is assessed by comparing the performance of experiments using the multi-angle information to the base case of a single WorldView-2 image classified using the multispectral information.

The multi-angle classification problem is introduced in Section 6.1. The remaining three sections present twenty total classification experiments from three parallel studies that separately investigate different aspects of the multi-angle information. Each study is conducted using unique land-cover classes that seek to emphasize the characteristics of each data set. These results, as well as the results from the model portability experiments, are concluded in Chapter 7.

6.1   Intro: Land-Use/Land-Cover Classification

These experiments explore the information available in two WorldView-2 multi-angle data sets. Specifically, the ability to extract contextual, height, and multi-angle spectral information at very-high spatial resolution and use this information to improve urban classification is studied. The study is carried out with specific focus on the extraction and use of multi-angle spectral information across both in-track and composite multi-angle data.

This dissertation uses data collected by WorldView-2 with eight spectral bands of information. The eight bands of spectral information is collected at very-high spatial resolution (2 m multispectral and 50 cm panchromatic). In addition, the operation of this type of instrumentation from a satellite platform provides the capability to accurately and repeatedly measure large urban areas that are the focus of these studies.

Exploring the ability to add classification accuracy to the foundation of this multispectral information, three classification studies are carried out:

(1) In-track multi-angle derived surface height and VHR contextual information (Section 6.2).

(2) In-track multi-angle multispectral and feature extraction information (Section 6.3).

(3) Composite multi-angle combined surface height and multispectral information (Section 6.4).

6.1.1   Multi-Angle Information

The data available from multi-angle image sets provide information sources not available from a single observation. Various components of the optical multi-angle imagery can be described as separate information sources:

- BRDF: Multi-angle image sets contain direct observations of bidirectional reflectance distribution function (BRDF). This angular reflectance variability is aproperty of the individual surface types.

- Solar observational cross section: Objects with pitched surfaces, such as trees and some building roofs, will present a different observational cross-section at each angle. This is often described as a component of the surface BRDF.

- Spectral Angle Variability: Surfaces with varying reflectance in both time and angle can be described by an error term that encapsulates the spectral variation of a pixel through the multi-angle sequence.

- Sun glint minimization: The effects of sun glint are minimized. When a surface is obscured by a solar specular reflection in one image due to the geometry of the multi-angle collection, the reflectance will be intact in the remainder of the sequence and available for the classification system.

- Pixel Height (Derived Spatial Correlation): Pixel height provides an explicit value for height above the ground derived by subtracting the elevation model (bare earth) from the surface model, yielding a terrain normalized height value for objects in the scene.

- Seasonality: In the case of composite multi-angle data sets, the changes in season can provide an additional signal for class types with seasonal dependency, such as vegetation.

These information sources can play an important part in identifying land-use classes. The aim of the following experiments is to extract and leverage these information sources from multi-angle imagery.

The utility of these parameters is, of course, class dependent. However, the unique characteristics of some land-use classes that are typically difficult to identify can be described by one or more of the multi-angle information sources. For instance, the inclusion of pixel height information increases the ability to differentiate objects with significant height compared to other spectrally similar classes, such as bridges and highways, buildings and skyscrapers, as well as trees and ground level vegetation. BRDF signals are often associated with vegetation classes, though can also be present in man-made surfaces providing a means of differentiation between spectrally similar classes such as trees and grass. The spectral variability can be indicative of motion through a scene, and, therefore, descriptive of land-use classes such as highways where the movement of vehicles significantly impacts the spectral signature of the asphalt.

6.1.2   Multi-Angle Difficulties

The development of multi-angle data for urban land-use classification requires the investigation of several open questions regarding the use of multi-angle data. In this section, these difficulties

are outlined. Since these approaches have not been extensively explored, each issue constitutes a necessary area of research for this project. These are treated as secondary objectives for this project and are investigated to the extent necessary to create a well defined approach for multi-angle urban classification.

Each of the items listed below arise through the inclusion of multiple views of a single scene into a machine learning system. In all cases, these issues are unique to multi-angle imagery and constitute a novel area of research as applied to urban classification:

(1) View-angle Spectral Distortion: The atmospheric effects inherent in multi-angle observations are significant. Since each observation is taken at a different angle, the spectral observations recorded at the satellite vary significantly due to the increasing atmospheric optical depth in off-nadir imagery. Due to the unique nature of multi-angle classification, spectral data space normalization could play a significant role in increase machine learning accuracy with multi-angle imagery.

- In this dissertation, the spectral distortion associated with satellite view angle is removed from the imagery through atmospheric compensation that accounts for the atmospheric conditions at the time of observation as well as look angle of the satellite.

(2) Image Registration: In order to extract information from the full multi-angle image stack, it is necessary to query the same pixel from each scene. This generally requires co-registered images in the case of moderate or low resolution satellite imagery. However, in urban very-high spatial resolution imagery, pixel selection also requires correction for topographic effects due to observation angle, i.e. building lean.

- In this dissertation, each of the images in the multi-angle data sets are created as true-orthos and bundle adjusted to the other images in the sequence. The true-ortho imagery uses a very-high resolution digital surface model to accurately correct the distortion in the imagery due to vertical structure. This creates a sequence of images in which every object is in the same image location, regardless of view-angle and

vertical height. By bundle adjusting these images, each object can be directly queried in all imagery by pixel location.

(3) Hughes Phenomenon: Multi-angle data sets inherently have a large amount of information. The dimensionality of the full data set is the number of spectral bands multiplied by the number of images included in the multi-angle analysis. This additional information fundamentally increases the complexity of the machine learning problem, and, therefore, decreases the accuracy of the classification. This is known as the Hughes phenomenon and is more commonly referred to as the "curse of dimensionality."

- In this dissertation, the large dimensionality of the multi-angle data space is addressed using both statistical and physical feature extraction methods. The feature extraction techniques reduce the dimensionality of the data set by reducing the data space to a smaller set of features that more concisely describe the data. Statistical methods, such as principal components analysis, use the structure of the data to determine an appropriate decomposition of the data space into feature that carry the most information. Physical methods use knowledge of the data space, such as the angularity of the spectral observations, to decompose the data space into fewer physically descriptive features than that of the original data space.

(4) Seasonality: Composite multi-angle data sets can include data from multiple seasons. This is generally considered a source of confusion in machine learning. There are methods for leveraging dense seasonal information as a signal source [28], however, there is no guarantee of a well sampled seasonal signal in multi-angle data sets.

- In this dissertation, seasonality is leveraged as a means to improve classification performance through statistical extraction of the temporal variations of LU/LC classes.

6.2   In-Track Multi-Angle LU/LC Classification with Spatial Information

Eight experiments based on the 50 cm spatial resolution panchromatic data investigate the classification contribution of multi-angle derived height information, in addition to the contributions of texture and morphological features. The multispectral information is incorporated in these experiments at the 50 cm spatial resolution through the use of pan-sharpening [81].

This multi-angle analysis is presented as two separate studies. The first study, presented in this section, focuses on the spatial information available from the full image sequence. This information includes the pan-sharpened spectral data, texture and morphology profiles based on the panchromatic data, as well as the multi-angle derived pixel height information. The development of these data sets is discussed further below. The second study, presented in Section 6.3, will focus on the spectral information in the multi-angle sequence.

Between the two studies, thirteen classification experiments are conducted. Experiments one through eight are part of the spatial analysis. Each of these classification experiments are conducted using the Random Forest algorithm [17] with the inputs shown in Table 6.1. The Random Forest is trained with ten trees and no limits on depth or features.

The feature contribution is based on the information gain ratio [90]. The final feature contribution is a normalized accumulation of information gain ratios for each of the classes in the analysis. For each class, the information gain ratio is calculated for the individual class against the combined remaining classes. This provides a feature contribution measurement for each input (as measured against the remaining classes) and by class (the calculated gain ratio). The information can then be combined to present feature contribution grouped by class (Figures 6.5 and 6.10) or input data set (Figures 6.6 and 6.11). By conducting these calculations separate from the classification algorithm, this approach remains classifier independent and is not limited to analyses that employee decision tree type learners.

The balanced F-measure (also F_1 score) [78, 106] is calculated in each experiment for all classes. This metric is used in an effort to balance the precision and recall of each class and is calculated as $F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$.

6.2.1   Methods: In-Track Spatial & DHM Classification

Three data sets of spatial measurements are developed in addition to the pan-sharpened [81] base case of a single nadir image. These three data sets are measurements of texture, morphology, and height. With these three derived data sets, and the nadir base case, eight experiments are devised, as shown in Table 6.1. These experiments are designed to explore the classification improvements available by including the additional information on top of the base case of a single pan-sharpened image (Exp. #1).

The texture data set is created using six second-order textural parameters calculated using the panchromatic band of the most nadir image [41, 42]. The six texture parameters are: homogeneity, contrast, dissimilarity, entropy, second moment, and correlation [80]. The texture parameters are calculated over three window sizes, 5x5, 15x15, and 51x51 pixels, which cover the range of features typically found in an urban area [96]. Only the diagonal direction with a cell shift of 15 pixels was chosen for these calculations. This creates a total of 18 input features for the textural component of the classification experiments.

The morphological data set is created with opening and closing by reconstruction [24]. Six profile components are calculated using the two morphological operators over the same three window sizes (5x5, 15x15, and 51x51 pixels) used for the texture calculation [104]. This creates a total of six morphological classification input features.

The pixel height information is contained in a digital height map (DHM). This is a map of height above the surface for each pixel in the image. The DHM (Figure 6.1c) is calculated by removing the digital elevation model (DEM) (Figure 6.1b) from the digital surface model (DSM) (Figure 6.1a) calculated for the study area. These are created using BAE Systems' Socket Set NGATE module [31]. The approach uses the known satellite observation angles and spatial cor-

Table 6.1: Spatial data set inputs for experiments one through eight and the number of features included in each data set.

Data Set	Features	Exp. #1	Exp. #2	Exp. #3	Exp. #4	Exp. #5	Exp. #6	Exp. #7	Exp. #8
Pan-sharp	8	X	X	X	X	X	X	X	X
Texture	18		X			X	X		X
Morphology	6			X		X		X	X
DHM	1				X		X	X	X

relation between the thirteen multi-angle images to extract a height measurement for each pixel in the scene. These products are computed with no human pre- or post-processing to keep the development as hands-off as possible.

As evident by the height maps and renderings in Figure 6.1, the resulting DHM shows significant distinction between skyscrapers and other lower lying classes, such as commercial building, residential homes, and trees. In turn, these lower lying classes can also be distinguished from ground level classes, such as roadways and grass. Figures 6.1d, 6.1e, and 6.1f show three dimensional renderings of the Atlanta area illustrating the accurate reconstruction of buildings, urban canyons, and highway bridge structures from the DHM.

From the pan-sharpened image, 15 ground survey classes are selected as shown in Table 6.2. These represent a wide variety of both natural and man-made land cover and include bridges, parking lots, vehicles, and skyscrapers. About 2.6 million pixels were collected by careful visual inspection for validation purposes, whereas 1,000 independent samples per class are used for training.

6.2.2   Results: In-Track Spatial & DHM Classification

Inclusion of the spatial data affects the classification results in two ways. First, the inclusion of the DHM information bolsters the ability to classify anything with a significant height compared to
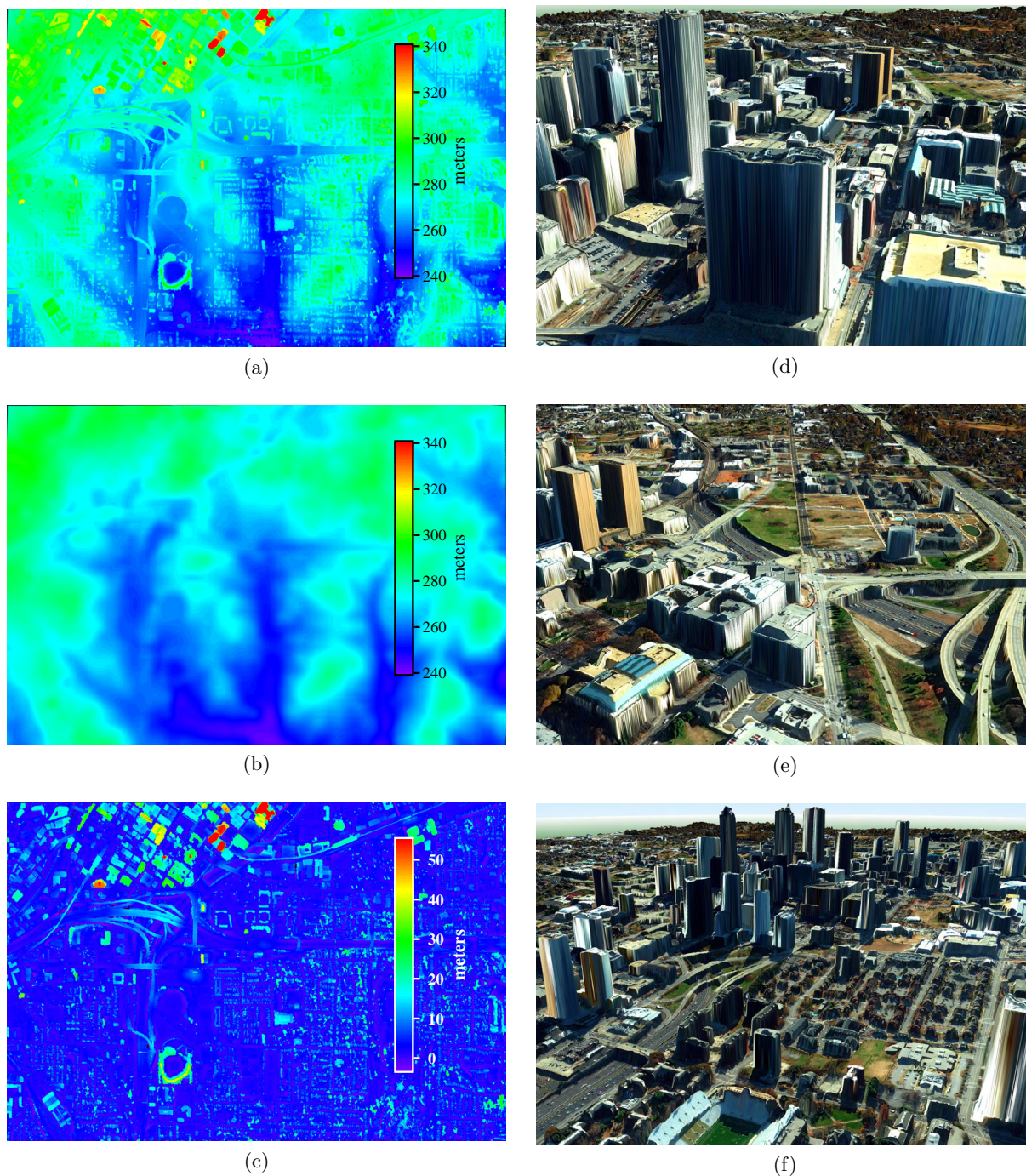
Figure 6.1: The DSM (a) and DEM (b) calculated from the multi-shot sequence with values measured in meters above ellipsoid (HAE). Subtracting the DEM from the DSM produces the DHM (c) with height measured in meters above the surface. Three images near downtown Atlanta are draped over the DSM in (d), (e), and (f) as an illustration of the multi-angle extraction of buildings, urban canyons, and other urban features.
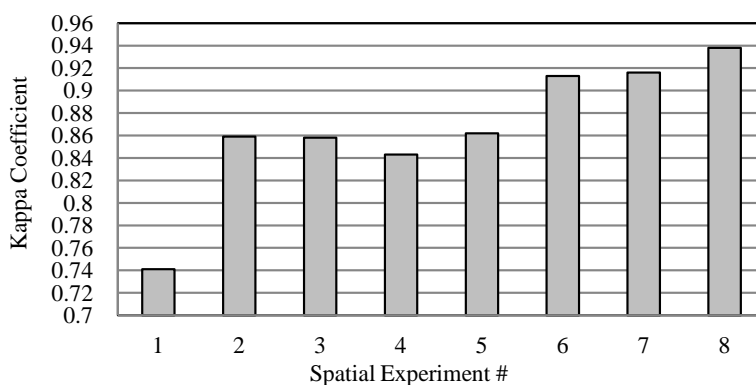
Figure 6.2: Kappa coefficient of the eight spatial experiments. The data inputs for each experiment are shown in Table 6.1.

other spectrally similar classes, such as bridges and highways. In addition, texture and morphology provide contextual awareness not available on a pixel-by-pixel basis. This allows the classifier to take into account spatial information when classifying a pixel that is part of a larger object.

Across the eight experiments, there is a steady improvement over the base case (Exp. #1). Figure 6.2 shows that the inclusion of the derived spatial data improves the kappa coefficient in all cases over the pan-sharpened experiment. The base case of the pan-sharpened image has a kappa coefficient of 0.74. The improvement in the most sophisticated case of experiment eight, utilizing the pan-sharpened, morphology, and texture information, in addition to the DHM, produces a kappa coefficient of 0.94. This represents a 27% improvement over the base case. Figure 6.3 shows the classification map of this experiment.

Figure 6.4 shows a boxplot of the 15 spatial classes illustrating the mean value and distribution of the height data in each class. As expected, the three structure classes, residential home, building, and skyscraper, are well identified. The tree class shows a taller height distribution when compared against the spectrally similar stressed and healthy vegetation classes. The bridge class is well separated from the spectrally similar highway and residential roadway classes. The shadow class is an anomaly and has no defined height structure as shadows occur both on the ground and on building tops.
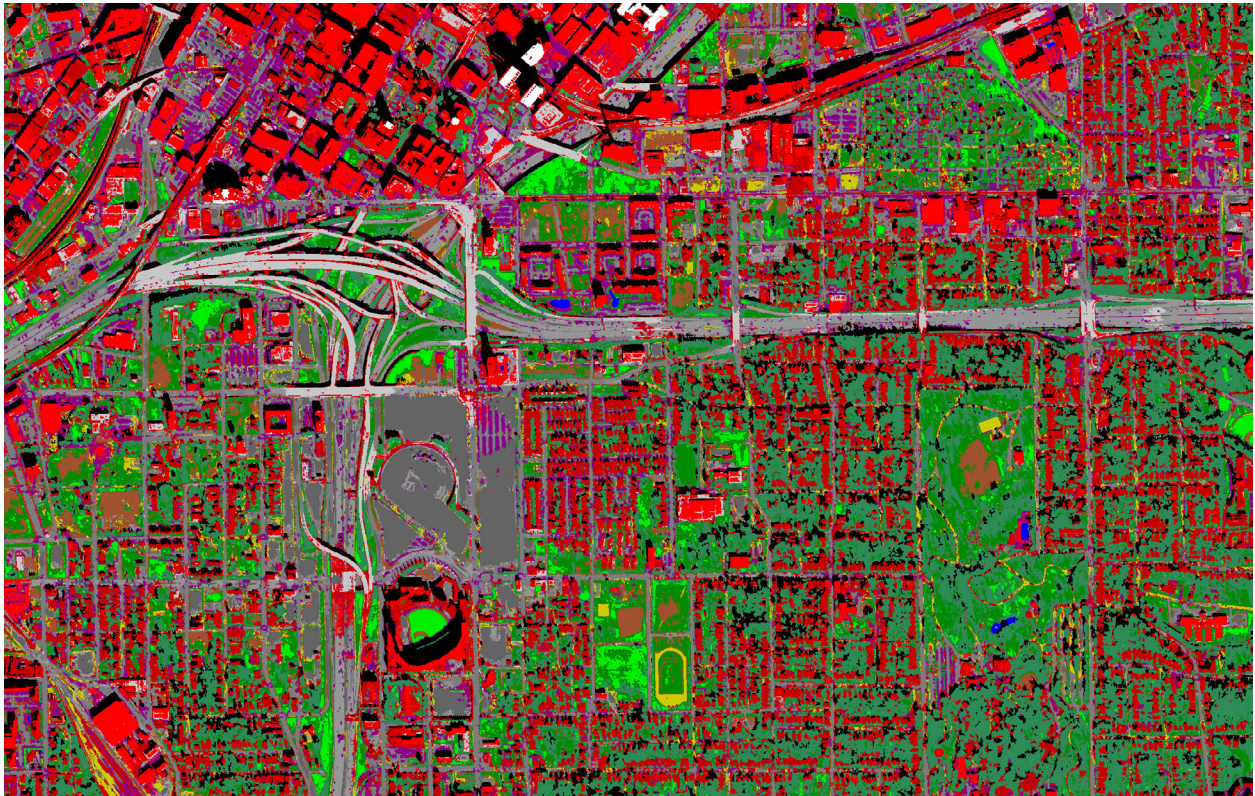
Figure 6.3: Classified image of the Atlanta, Georgia scene from spatial experiment eight. Color legend: ☐ = Skyscraper, ■ = Building, ■ = Residential Home, ☐ = Bridge, ■ = Highway, ■ = Residential Road, ■ = Parking Lot, ■ = Healthy Vegetation, ■ = Stressed Vegetation, ■ = Soil, ■ = Trees, ■ = Vehicles, ■ = Recreational, ■ = Shadow, ■ = Water.
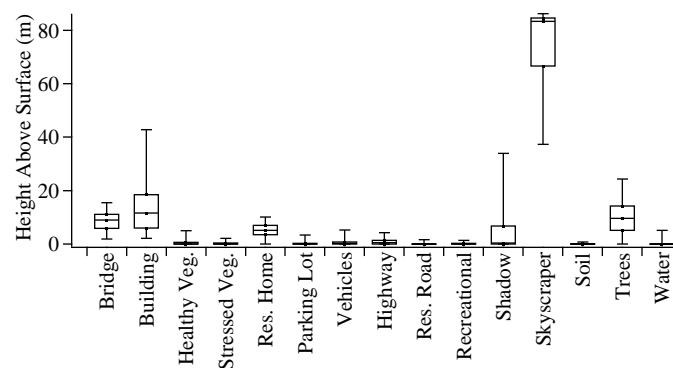
Figure 6.4: Boxplot of the ground survey DHM values for the 15 spatial classes (whiskers plotted to 2nd and 98th percentile, outliers omitted).

Table 6.2: F-measure (percentage) for each of the fifteen land cover types used in the eight spatial experiments and the average F-measure for each experiment. The method with the highest F-measure is highlighted for each class. The data set inputs for experiments #1 through #8 are shown in Table 6.1.

Class Name	Exp. #1	Exp. #2	Exp. #3	Exp. #4	Exp. #5	Exp. #6	Exp. #7	Exp. #8
Skyscraper	45	48	57	96	54	74	94	77
Building	75	84	86	89	81	92	94	94
Res. Home	45	71	73	65	75	80	83	85
Bridge	72	79	86	85	84	89	95	94
Highway	36	71	77	60	82	80	84	88
Res. Road	69	84	78	79	84	91	85	93
Parking Lot	83	93	91	89	93	96	93	97
Healthy Veg.	95	98	98	97	99	98	98	99
Stressed Veg.	95	96	95	96	96	96	97	97
Soil	91	94	95	94	96	95	96	96
Trees	88	94	93	93	95	95	95	96
Vehicles	25	81	63	53	83	91	73	93
Recreational	84	85	89	90	92	88	90	90
Shadow	92	97	96	93	95	97	97	98
Water	74	83	83	78	56	81	86	82
Average F-measure	71	84	84	84	84	90	91	92

Table 6.2 shows classification F-measure for the 15 spatial classes in experiments one through eight. The feature contribution is presented in two graphs. The feature contribution, broken down by class, is shown in Figure 6.5. The same information is presented by class and broken down by information group in Figure 6.6.

A significant amount of information can be extracted from these two feature contribution plots. They show that many of the spectral, textural, and morphological inputs provide significant feature contribution, but that the DHM contributes most significantly to the overall classification.
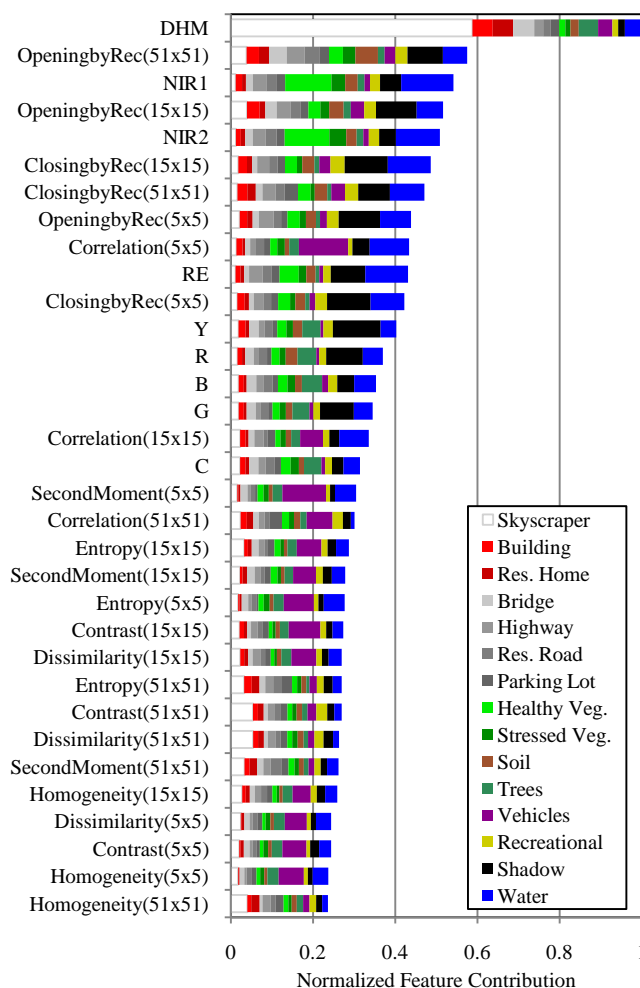
Figure 6.5: Feature contribution, grouped by class, for the full spatial experiment incorporating all input data (Exp. #8).

The skyscraper, building, residential home, and bridge classes are most affected by the DHM information. Conversely, the small-scale texture information is the most important contributor to discrimination of the vehicle class. This can be seen in both feature contribution plots. Each texture input in Figure 6.5 shows a significant contribution to the vehicle class (purple) compared to the other classes. Figure 6.6 shows the most significant contribution to the vehicle class came from the combined texture inputs.

Of particular interest is the skyscraper class. Experiment number four (pan-sharpened and DHM inputs) shows the highest level of F-measure for this class. When the textural and morpho-
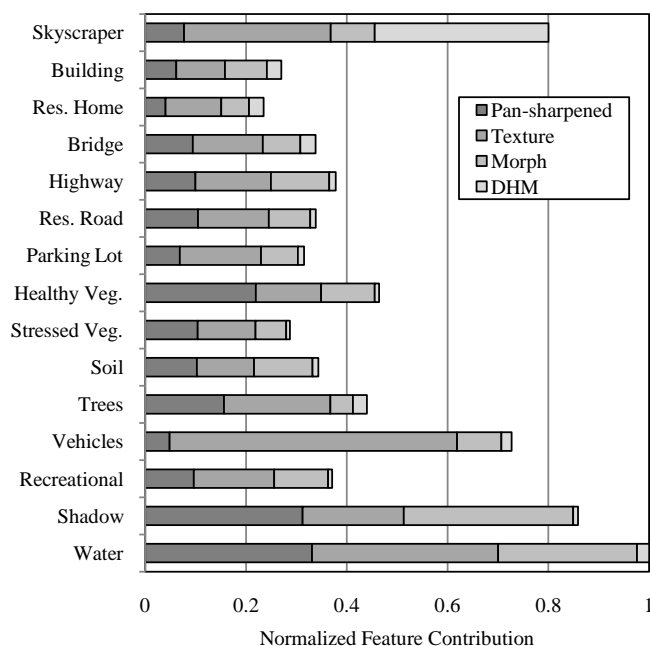
Figure 6.6: Feature contribution, grouped by data type, for the full spatial experiment incorporating all input data (Exp. #8).

logical information are added to the analysis, the skyscraper class F-measure declines significantly. Figure 6.6 shows that the feature contribution of the texture information, with 13 features, is of approximately the same level as the single DHM input when including all features (Exp. #8). This indicates a conflict between the discrimination of the skyscraper class due to the height information of the DHM and the surface material information of the texture measurements.

It is also interesting to note that the pan-sharpened information accurately classified healthy vegetation. Table 6.2 shows very little improvement in the F-measure performance for this class over the base pan-sharpened case. This can be seen in Figure 6.5 where the largest contributor to the healthy vegetation class (bright green) is the pan-sharpened information, especially the red edge and the two near infrared bands. Figure 6.6 shows that the pan-sharpened information contributed most significantly to the classification when compared against the additional contribution from the texture, morphological, and height information.

Figure 6.7 shows three examples from the classification map. In these examples, the classified image shown is from the eighth experiment with the highest kappa coefficient.

Figure 6.7a shows a building with roof-top parking. Figure 6.7d shows the corresponding classified image. With spectral information alone this was classified as a parking lot. However, as evidenced by the shadow, it is a building with roof top parking. In addition, the classifier accurately identified vehicles in the roof top parking lot and on the ground below.

Figures 6.7b and 6.7e show the discrimination between bridges and highway grade. The transition between the two classes is generally evident on the roadway despite some false positives of the building class near the large elevation gradients between the bridge and ground level. The small commercial building, parking lots, and vehicles, in the bottom right of the image chip, are also accurately delineated.

Figures 6.7c and 6.7f show a further example of parking lot and vehicle classification. In the upper right of the image, there is a large parking lot where nearly all vehicles are identified. The asphalt in proximity to the vehicles is classified as a residential roadway, and the general parking area, where traffic is less prominent, is accurately delineated as a parking lot. Additionally, the baseball practice area in the bottom left of Figures 6.7c and 6.7f is possibly better identified by the classification system than by eye. There is a fence line that follows the outfield of the paved practice field and is barely visible in the image. However, with the additional help of the DHM, it is accurately classified as a small structure (red) in Figure 6.7f.
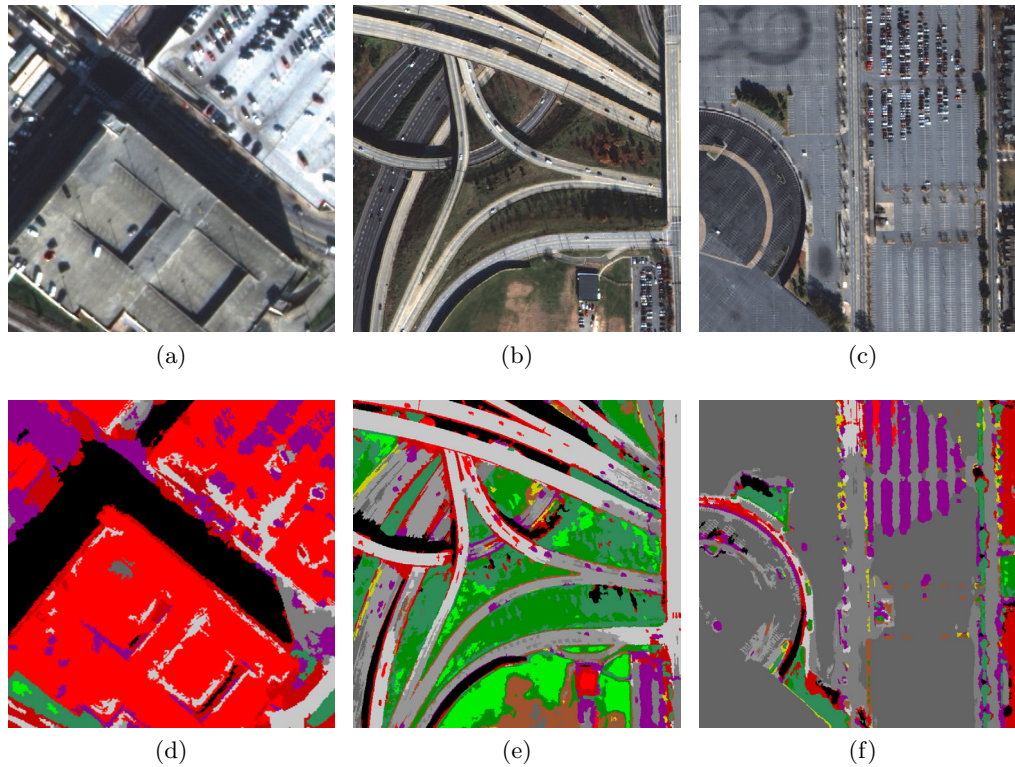
Figure 6.7: Details from the eighth spatial experiment for roof top parking (a)/(d), bridges (b)/(e), and a parking lot with baseball practice area (c)/(f) (applicable color legend is in Figure 6.3).

6.3   In-Track Multi-Angle LU/LC Classification with Spectral Information

Five experiments are conducted using the 2 m spatial resolution multi-angle multispectral data. These experiments explore the contribution of MAR using the eight band multi-angle multispectral data set as well as three feature extraction data sets (reducing the large size of the raw data space).

The second study in this paper, presented in this section, investigates the spectral information. Experiments nine through thirteen are used to study the classification impact of the multi-angle multispectral information. The basis of this analysis is the thirteen WorldView-2 images described in Section 3. These images have two-meter spatial resolution with data in eight spectral bands. Additional information is extracted from this data set using three feature extraction techniques. The first is a physically-based feature extraction that models the observed spectral

change across view-angle. The second technique employed an unsupervised principal component analysis (PCA) [73] while the third uses the supervised non-parametric weighted feature extraction (NWFE) method [56, 57]. The development of these data sets is further discussed below.

Each of these classification experiments are conducted using the Random Forest algorithm, information gain ratio feature contribution calculations, and F-measure calculation, as described in Section 6.2.

6.3.1 Methodology: In-Track Spectral Classification

The first spectral experiment (Exp. #9) uses only the most nadir multispectral image. These experiments seek to explore the additional capabilities available from multi-shot sequences and, therefore, this case is treated as the spectral performance baseline.

Experiment ten uses a data set composed of the full stack of 13 multispectral images. This is a very large amount of data, totaling 104 features. Classification with this large amount of raw data will most likely be less successful than other techniques due to the curse of dimensionality [10, 15]. However, this is an important case for this study as it is the most direct classification method utilizing the full range of spectral data available.

Experiment eleven uses a data set developed from a novel physically-based feature extraction technique. In this technique, a polynomial fit is used to characterize the MAR curve of the multi-angle sequence. For each pixel in a spectral band, a second order polynomial curve is fit through the spectral signature as it changes across view-angle. These polynomial fits are of the standard form ($y = ax^2 + bx + c$), where the x-axis is the off-nadir view-angle and the y-axis is the pixel reflectance. The coefficients and error term represent:

a) magnitude of the spectral data curve (x^2)

b) linear trend of the spectral data (x)

c) spectral offset of the multi-angle fit (y-offset)

d) standard error of the fit (y-fit error)

Table 6.3: Spectral data set inputs for experiments nine through thirteen as well as the number of data features used in each experiment. The experiments are ordered by the number of features in each data set, with the exception of the baseline experiment (Exp. #9).

Imagery Data Sets		
Nadir Multispectral	Exp. #9	8 features
Multi-angle Multispectral	Exp. #10	104 features
Feature Extraction Data Sets		
Polynomial Multispectral	Exp. #11	32 features
Principal Component Analysis	Exp. #12	16 features
Nonparametric Weighted	Exp. #13	11 features

In this way, four physically meaningful components are extracted from the full data set. The components mainly correspond to the information sources discussed in Section 6.1: partial BRDF (x^2), solar observed cross section (x), and pixel reflectance variability (y-fit error) as well as surface reflectance (y-offset).

For the data set used in experiment eleven, these parameters populate four layers for each of the eight WorldView-2 spectral bands. This results in a total of 32 features describing the spectral variation of each pixel through the multi-angle sequence.

While this reduction of dimensionality is a significant improvement over the 104 multispectral features of the full multi-angle sequence, it is still a relatively large data set. To further reduce the dimensionality of the data set, two additional feature extraction techniques are implemented.

The first is the classic principal component analysis feature extraction [73] (Exp. #12) applied to the full image stack of experiment ten. This results in 16 features that contain 99% of the full image stack variance. This data set is used as input for spectral experiment twelve.

The remaining data set (Exp. #13) is extracted from the full image sequence using the non-parametric weighted supervised feature extraction technique [56] [57]. Feature depth is optimized through the use of accuracy testing across the full 104 features of the transformed data space. This results in a data set containing 11 total features used for classification input in experiment thirteen.

Generating a ground survey from a multi-angle dataset presents a unique set of challenges that revolve around the ability to accurately extract spectral information for a specific target in each of the multi-angle images. Generally, there are three possible methods for creating a ground survey that encompass all images in the multi-angle sequence:

(1) An individual ground survey can be manually selected from each image. This would be excessively time consuming and could produce an unbalanced data set across the image sequence.

(2) A single image ground survey can be propagated to all multi-angle images through the use of a sensor model able to rigorously propagate the ground coordinates from image to image. This process would be very accurate, but require significant development and validation to appropriately deploy.

(3) A single image ground survey can be propagated to all multi-angle images through the creation of bundle-adjusted true-orthorectified images from the collected imagery and the DSM calculated as part of the spatial investigation.

In these experiments, the third approach is used. This provides the means to directly select the corresponding pixel in every image of the sequence. However, standard true ortho processes produce artifacts in areas where building lean is removed. This creates completely unusable information, by necessity, in areas where the sensor has no data due to observation angle (occlusion). In an urban scene, this is a relatively large effect. This necessitates that the ground survey be isolated from rapid changes in elevation that could produce processing artifacts at some point in the image sequence. This is to say that all samples are collected in open areas of relatively small vertical change, such as on building tops, in an open fields, or in the middle of roadways. This ensures that, for any given pixel selected, data is present in all the observation angles and ground survey data are not contaminated by artifacts of the true ortho processing.

For these spectral experiments, a ground survey different from the one used for the spatial experiments is created to better emphasize the MAR capabilities. The number of validation pixels
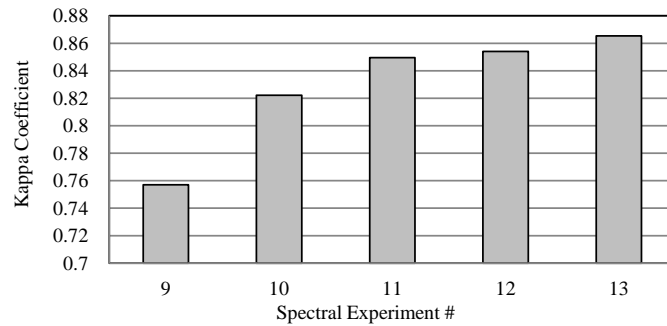
Figure 6.8: Kappa coefficients of the spectral experiments. The data inputs for each experiment are shown in Table 6.3.

in the spectral experiments is significantly smaller than that of the spatial experiments. It is necessarily limited by the lower spatial resolution of the multispectral imagery as well as the practical constraints of working with a multi-angle data sets and generating a ground survey that, for each pixel, ensures that the full MAR is represented.

Table 6.4 shows the fifteen class types selected for the spectral experiments. These represent a wide variety of both natural and man-made surfaces such as flat and pitched roofs, parked cars, and multiple vegetation types. One hundred independent pixels per class are used for training. Approximately ninety thousand ground survey pixels are selected by careful visual inspection for validation.

6.3.2   Results: In-Track Spectral Classification

Classification of the urban scene using MAR shows significant improvement over the base case of a single nadir multispectral image. Figure 6.8 shows the kappa coefficients from these five experiments. All methods that incorporate the MAR show improvement over the base case of a single nadir multispectral image. Furthermore, the three feature extraction techniques show additionally improved classification performance. The full classification map of the polynomial feature extraction experiment is shown in Figure 6.9.

The base comparison case of a single WorldView-2 image is used in experiment #9 and the full stack of imagery, with no feature extraction, is used in experiment #10. The performance
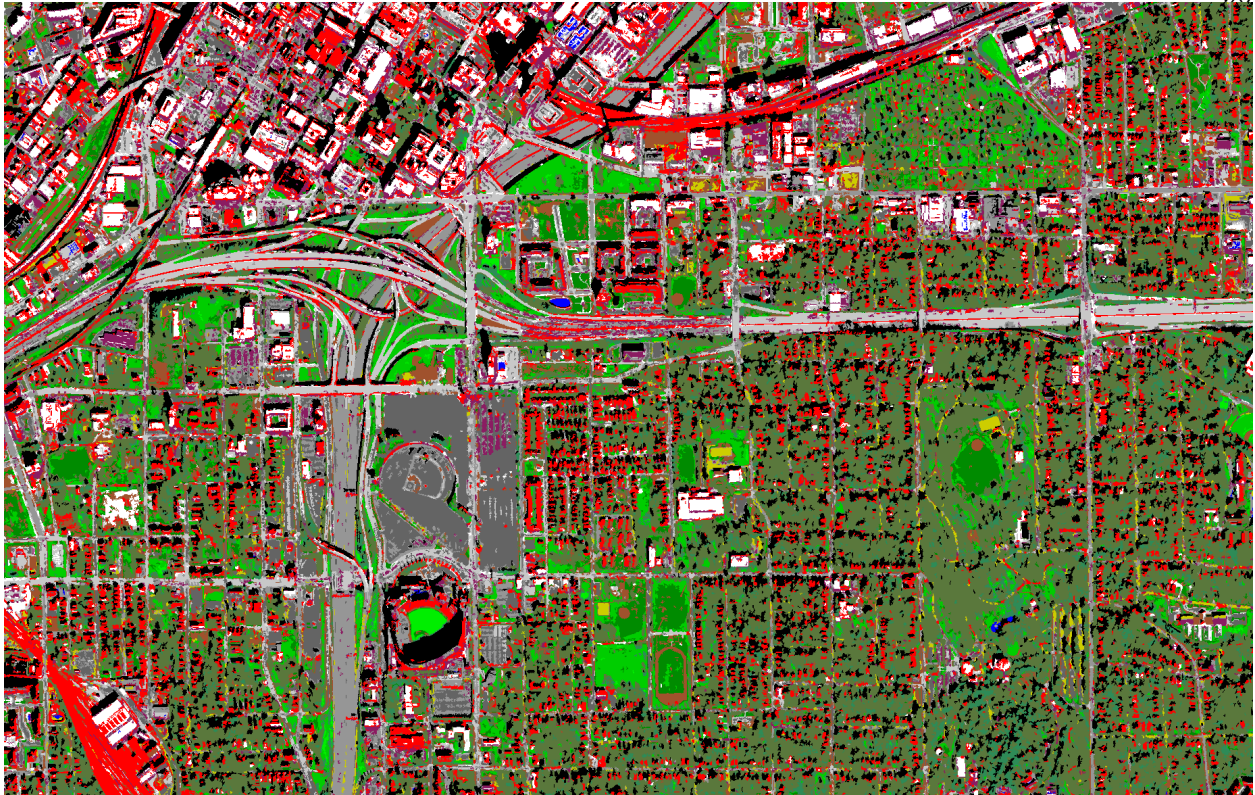
Figure 6.9: Classified image of the Atlanta, Georgia scene from the MAR polynomial fit experiment. Color legend: ☐ = Flat Roof, ■ = Pitched Roof, ☐ = Concrete Roadway, ■ = Asphalt Roadway, ■ = Parking Lot, ■ = Healthy Vegetation, ■ = Stressed Vegetation, ■ = Dormant Vegetation, ■ = Soil, ■ = Evergreen Trees, ■ = Deciduous Trees, ■ = Parked Cars, ■ = Recreational, ■ = Shadow, ■ = Water.

of these experiments, as shown in Figure 6.8, demonstrates that the multi-angle data provides an improvement of approximately 9% in it's most raw form. The feature extraction techniques used in experiments #11 - #13 provide an additional improvement on the order of 5%. The resulting classification maps produced by all three techniques are statistically different from each other, based on the McNemar's test [34] (to the 95% confidence level), and prove increasingly accurate as the data dimensionality decreases.

The polynomial feature extraction technique provides the ability to reduce the full spectral multi-angle data set into physical components describing the BRDF, observable solar cross section, reflectance, and spectral variability (see Section 6.1). This feature extraction technique yields the opportunity to study the importance of this information by class. This is a unique capability when
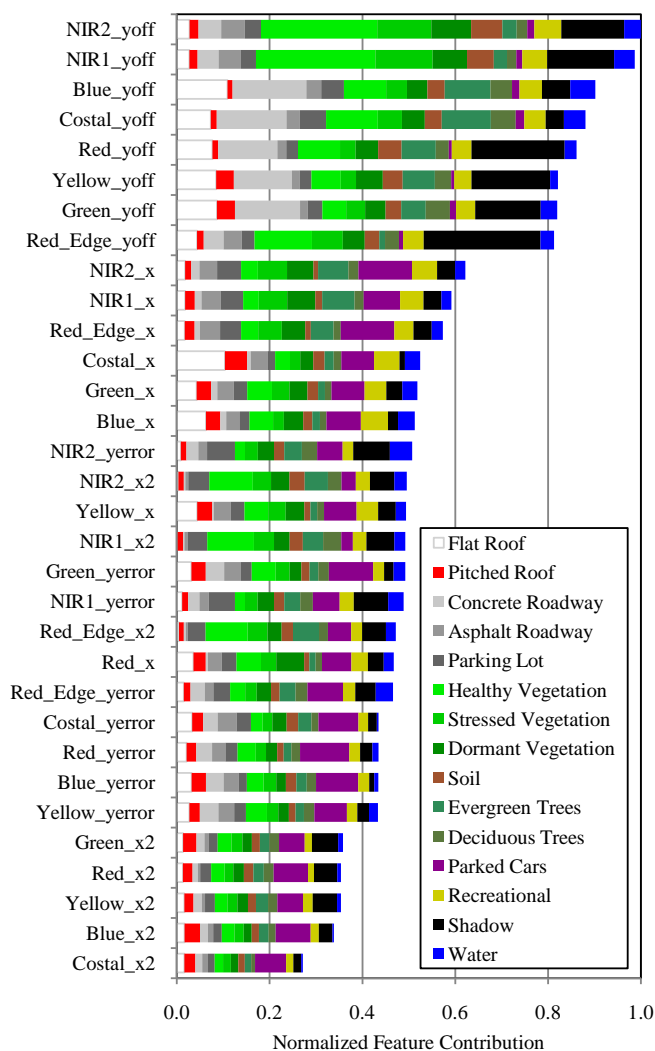
Figure 6.10: Feature contribution, grouped by class, for the spectral polynomial fit experiment (Exp. #11).

compared to the other feature extraction techniques used in this study, both of which provide an optimized data stack with no physical correlation. To this end, the feature contribution of experiment eleven (polynomial fit) is presented in two graphs. The feature contribution, broken down by class, is shown in Figure 6.10. The same information is presented by class and broken down by information group in Figure 6.11.

Figure 6.10 shows that the multispectral data, in this case encoded in the y-offset terms, are the most powerful of the inputs. However, the input contribution is class dependent. As shown
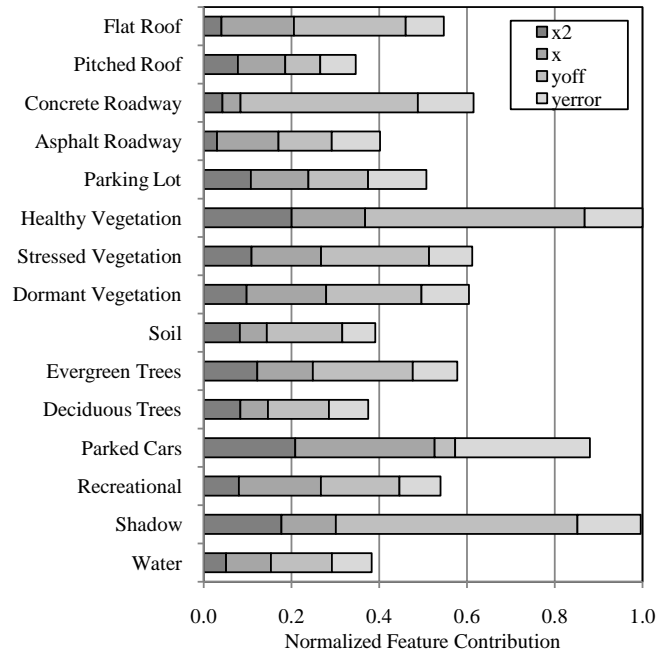
Figure 6.11: Feature contribution, grouped by data type, for the spectral polynomial fit experiment (Exp. #11).

in Figures 6.10 and 6.11, the vegetation classes have a high dependence on the red edge and near infrared spectral data. Conversely, the parked car and pitched roof classes have little dependence on the spectral information and instead rely on the standard error (y-fit error) and linear trend (x) terms, respectively.

Table 6.4 shows classification F-measure for the five spectral experiments and 15 classes. This allows for the comparison of individual class performance across the five spectral experiments. The experiment with the highest F-measure is highlighted for each class.

The five vegetation classes of the spectral experiments include healthy vegetation, stressed vegetation, dormant vegetation, evergreen trees, and deciduous trees. Across these classes, an improvement of 11% to 28% is seen, with the exception of stressed vegetation.

A possible explanation for the lack of improvement in the stressed vegetation class (with a maximum improvement of 3%) may be found in the spectral correlation with the deciduous class. Since the images were taken in winter time, deciduous trees have no leaves. Therefore, the deciduous class will, by necessity, include some underlying vegetation in the form of a spectral superposition

Table 6.4: F-measure (percentage) for each of the fifteen land cover types used in the five spectral experiments and the average F-measure for each experiment. The method with the highest F-measure is highlighted for each class. The data set inputs for experiments #9 through #13 is shown in Table 6.3.

Class Name	Exp. #9	Exp. #10	Exp. #11	Exp. #12	Exp. #13
Flat Roof	48	68	64	78	83
Pitched Roof	45	70	75	72	74
Concrete Roadway	87	92	91	94	93
Asphalt Roadway	69	79	84	85	82
Parking Lot	75	82	89	93	92
Healthy Vegetation	84	89	84	85	88
Stressed Vegetation	94	94	92	91	94
Dormant Vegetation	94	92	92	91	94
Soil	82	81	79	77	86
Evergreen Trees	73	80	85	71	80
Deciduous Trees	87	86	90	91	89
Parked Cars	24	43	43	41	44
Recreational	68	82	74	90	89
Shadow	89	92	89	89	91
Water	90	92	94	97	97
Average F-measure	74	82	82	83	85

with the vegetation classes observed through the tree canopy. This can be a particularly significant issue for the stressed vegetation class, which is spectrally similar to the leaf-off deciduous trees.

The vegetation class that most responded to the additional multi-angle and feature extraction information is the evergreen class. As opposed to the deciduous trees that are leaf off during the collection, evergreen trees can be identified by a changing multi-angle observed solar cross section, and therefore, a distinct polynomial linear fit term in experiment #11. It was in this experiment that a maximum F-measure of 85% is achieved for the evergreen class.

Additionally, three of the man-made materials classified in these experiments show a significant improvement available from the MAR information. The nadir base case shows a relative inability to classify the asphalt roadway and parked cars as well as the pitched and flat roof classes. These classes have an initial F-measure of 69%, 24%, 45%, and 48% respectively. The improvement for each of these classes, after adding the MAR information, is as high as 22%, 75%, 68%, and 70% respectively. The majority of these improvements are found in the polynomial feature extraction experiment and can be contributed to the extraction of physically meaningful parameters (as shown in Figure 6.10 and 6.11), such as pixel variability.

Figures 6.12a and 6.12d show an example of vegetation classification by the polynomial fit experiment. In the image, there are several evergreen and deciduous trees as well as a mix of stressed vegetation, dormant vegetation, and soil. Additionally, small flat and pitched roof structures are accurately identified in the image.

Figures 6.12b and 6.12e illustrate the polynomial fit classifier's capability to distinguish parked cars, parking lots, vegetation, residential houses, and trees. Despite some misclassification of the pitched roof and shadow classes near the top of the parking area, the classifier identifies the parked vehicles as well as many of the empty spaces in between. The vegetation is also generally well classified showing both stressed and dormant vegetation in the visible open field. Both deciduous and evergreen trees can be seen obscuring residential housing and bordering the parking lot.

Figures 6.12c and 6.12f illustrate the spatial distribution of temporal variation in the scene. These show the standard error of the polynomial fit (term d in Section 6.3.1) across a scene showing a roadway interchange, residential roads, and other urban features. Since each pixel in the scene is fit with a polynomial curve across the multi-angle sequence, the polynomial y-fit error term effectively encapsulates the temporal variation of each pixel during the satellite overpass. This can help distinguish between land-use types, such as low volume residential roads and high volume highways.
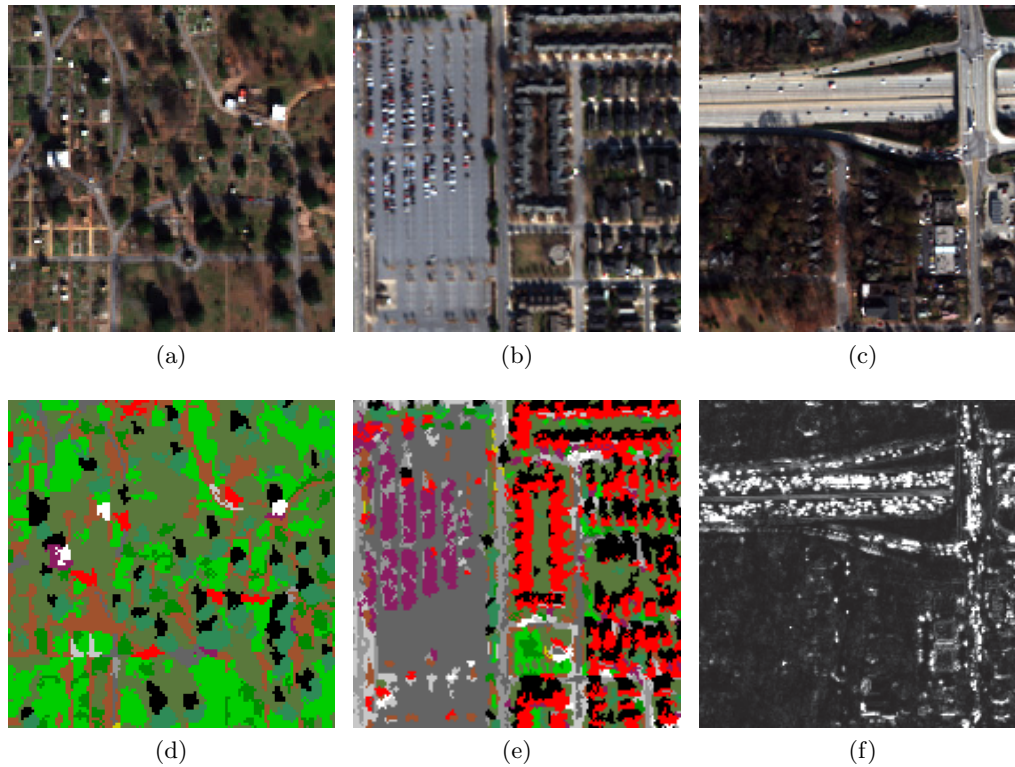
Figure 6.12: Polynomial y-fit error plot and classification examples of vegetation (a)/(d), a parking lot (b)/(e), and a roadway interchange (c)/(f) (applicable color legend is in Figure 6.9).

A boxplot of the pixel level y-fit error distribution for the 15 spectral classes is presented in Figure 6.13. It shows that the classes with the largest error signals are the two roadway classes and the parked car class. In these cases, the variation in reflectance of randomly oriented vehicle surfaces or movement of vehicles during the multi-angle collection causes the pixel MAR to vary significantly. These MAR pixels with poor polynomial fits produce a large y-fit standard error that can be used to map the pixels of high variability in an image.
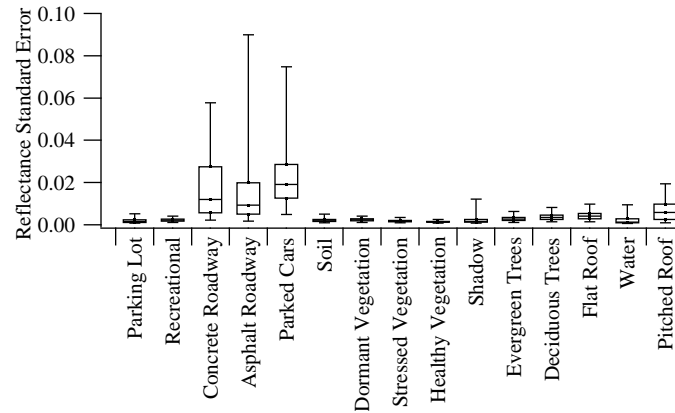
Figure 6.13: Boxplot of the green band y-fit error for the 15 spectral classes (whiskers plotted to 2nd and 98th percentile, outliers omitted).

6.4 Composite Multi-Angle LU/LC Classification

The section explores the ability to extend the lessons learned in the previous sections to the more generally available composite multi-angle data. In the previous sections, methods of utilizing the information available from in-track multi-angle data, including derived height and BRDF type information, have been explored. However, these in-track multi-angle data sets are relatively rare. In this section, similar techniques are applied to a more widely available composite multi-angle data set in an effort to explore multi-angle methods that may be applied on a large scale to improve land-cover classification.

The composite multi-angle data set that is the subject of these experiments was introduced in Section 3.2. It is a multi-temporal composite of 28 WorldView-2 images collection over the target area of Lakewood, CO over the course of 1.5 years. These images are well distributed over the azimuth and elevation observation hemisphere.

This data set is unique in that it provides multiple images of a single scene, collect over various orbits, from different view angles. This provides three unique data sources not present in a single WorldView-2 image:

(1) Surface reflectance properties/BRDF

(2) Seasonal spectral changes

(3) Derived pixel height

However, this additional information comes at the cost of added noise and processing complexity beyond that of the in-track cases discussed previously. These include different atmospheric composition between acquisitions, varying solar illumination, and real changes to the surface over time. The experiments in this section focus on extracting these information sources for the purposes of LU/LC classification.

6.4.1   Methods: Composite Spectral & DHM Classification

In the experiments that follow, the composite multi-angle data set is segmented in order to parametrically explore its temporal, angular, height, and combined information. In order to compare these different data sources, four data sets are created:

(1) Composite multi-angle data set (24 snow free images - Figure 6.14)

(2) In-track triplet multi-angle (3 images - Figure 6.15a)

(3) Multi-temporal nadir (3 images - Figure 6.15b)

(4) Optically derived pixel height

The full composite multi-angle data set described in Section 3.2 consists of 28 images. However, four of these images have significant snow cover and were, therefore, removed from the following experiments in order to reduce the noise sources and complexity of the experimental comparisons. Therefore, the full composite multi-angle data set used in these experiments contains the 24 images shown in Figure 6.14.

From these 24 images, an optically derived height map was created for the target area using the NGATE software discussed previously [31] for the in-track experiments. The height map is the pixel height above the surface. It is created by subtracting the digital surface model (optical surface) from the digital elevation model (terrain) over the entire target area. In this way, the height information can be included as an additional feature during classification.
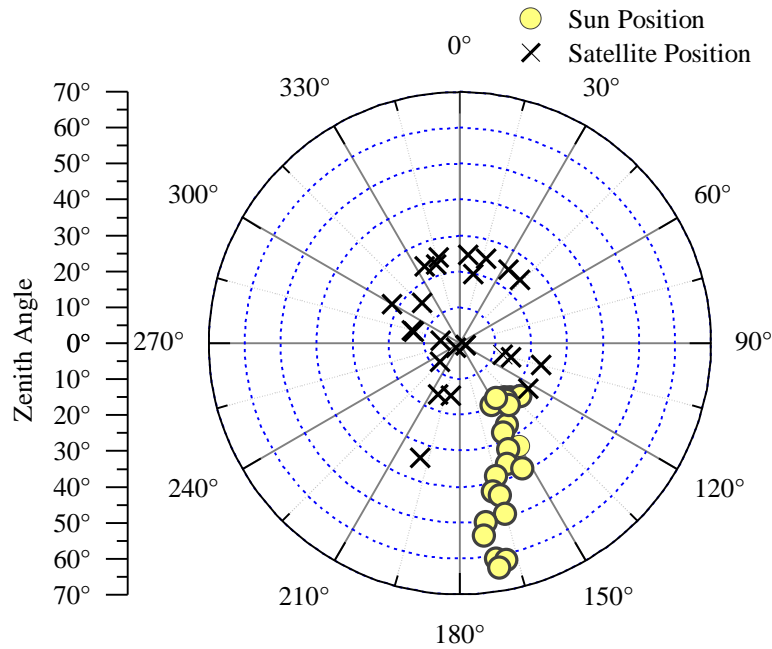
Figure 6.14: Ground observed azimuth and elevation of the satellite for each image observation (black marker) of the 24 image Lakewood data set. Azimuth is plotted angularly clockwise (north = 0°, east = 90°, south = 180°, west = 270°) and elevation is plotted radially from the center (ground nadir = 90°, ground horizon = 0°).

In addition, two three-image subsets are created from the 24 composite sequence in order to segment the multi-angle and multi-temporal components of the full data set.

The in-track multi-angle data is a triplet that was collected on July, 2001. This is a true in-track triplet, so the collection parameters are very similar to that of the previous section with the majority of changes in the sequence due to real BRDF type variations and observation angle atmospheric effects. In this data set, there is no multi-temporal information present, providing a good opportunity to parametrically explore the available classification contribution of the multi-angle information.

The nadir "triplet" is a combination of three images collected in May, July, and October in the middle of the composite multi-angle dataset. This is not a true triplet, in that the images were collected over multiple orbits, but all three of the images are collected with an off-nadir angle of less
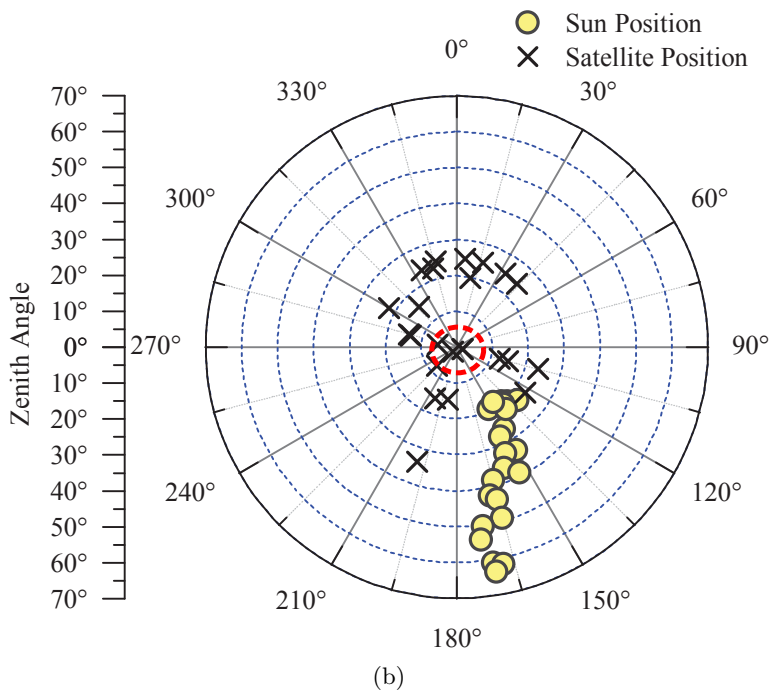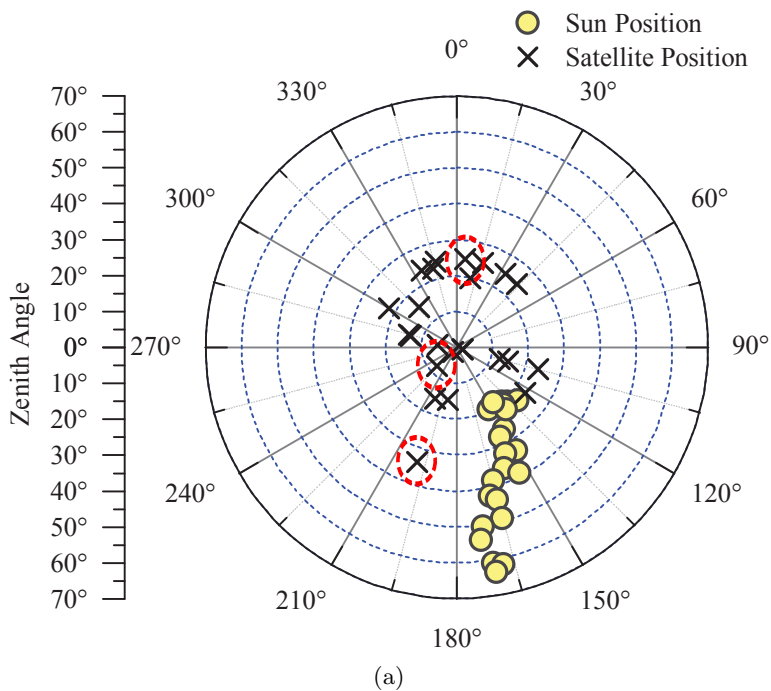
Figure 6.15: Azimuth and elevation locations of the in-track ((a)) and composite ((b)) triplet subsets used for the composite multi-angle classification experiments #15 through #18.

then 3°. This creates a dataset that segments the multi-temporal information from the multi-angle information of the in-track triplet.

In the previous multi-angle classification experiments, it was shown that information extraction is necessary when dealing with a large multi-angle data set in order to distill information from an otherwise unassociated stack of imagery. In the previous section, this was done by both creating an explicit physical model of the multi-angle space as well as through the use of statistical feature extraction techniques (PCA and NWFE). With these methods, data set features were created that represent the multi-angle BRDF-type information and provided the information source necessary to significantly improve classification accuracy over that of a single satellite observation.

In these experiments, the principal component analysis (PCA) method is used to the same effect. In each of the three data sets described above, the PCA is calculated and a feature stack containing 99% of the full image stack variance is extracted. The data sets are described, along with the previously discussed data, in Table 6.5. For the full composite multi-angle data set, this created a 20 feature data set that describes both the multi-angle and multi-temporal feature of the full data set. In addition, two six feature data sets were created from the multi-angle and multi-temporal triplets. In this way, a low-impedance method is used to create the extracted features necessary to describe the physical multi-temporal and multi-angle information present in the three data sets.

Using these data sets, seven experiments are performed. These are described in Table 6.5. In all but one of the experiments, the individual data sets are used in a Random Forest classification experiment in order to parametrically explore the information present in each. The single exception is the final experiment #20 where the DHM is added to the feature extraction full composite multi-angle data set in order to explore the maximum classification performance provided from incorporating all available data.

For the data sets used in each of the experiments, the same ground survey was extracted. This consists of 12 classes that cover an array over natural and man-made features: pitched roofs, flat roofs, asphalt roadways, parking lots, bridges, solar panels, skyscrapers, water, conifer trees,

Table 6.5: Feature stacks created from the composite multi-angle data set. Each of the feature stacks is referred to in Figures 6.17 and 6.19.

Imagery Data Sets		
Individual Multispectral	Exp. #14	8 features (24 images)
Three Nadir Stacked	Exp. #15	24 features
In-Track Triplet	Exp. #16	24 features
Feature Extraction Data Sets		
Three Nadir PCA	Exp. #17	6 features
Triplet PCA	Exp. #18	6 features
All Composite PCA	Exp. #19	20 features
Height Included Data Sets		
All Comp. PCA + DHM	Exp. #20	21 features

deciduous trees, natural grass, and maintained grass. A total of 40,000 pixels were selected from the data sets as shown in the RGB example of Figure 6.16. A total of 100 pixels per class is used for classification of the Random Forest with the remaining pixels reserved for validation purposes.

6.4.2   Results: Composite Spectral & DHM Classification

In the following section, the results from the seven composite multi-angle classification experiments are presented. The results show that the additional information available through this type of data set can improve classification accuracy by as much as 30% as measured by kappa coefficient. Additionally, this type of information allows the extraction of classes that are typically difficult to separate, such as pitched and flat roofs as well as asphalt roadways and parking lots.

The kappa coefficient from each of the experiments is shown in Figure 6.17. Each of the bars in the chart represents the kappa coefficient performance of a single experiment as described in Table 6.5, with the exception of the far left bar. This is the average single image classification performance for all 24 images in the composite multi-angle data set with the standard deviation of the images kappa coefficients represented by the error bars.

Figure 6.16: Ground survey of the Lakewood, CO scene. Color legend: ■ = Pitched Roof, ■ = Flat Roof, ■ = Asphalt Roadway, ■ = Parking Lot, ■ = Bridge, ■ = Solar Panels, ■ = Skyscrapers, ■ = Water, ■ = Conifer, ■ = Deciduous, ■ = Natural Grass, ■ = Maintained Grass.
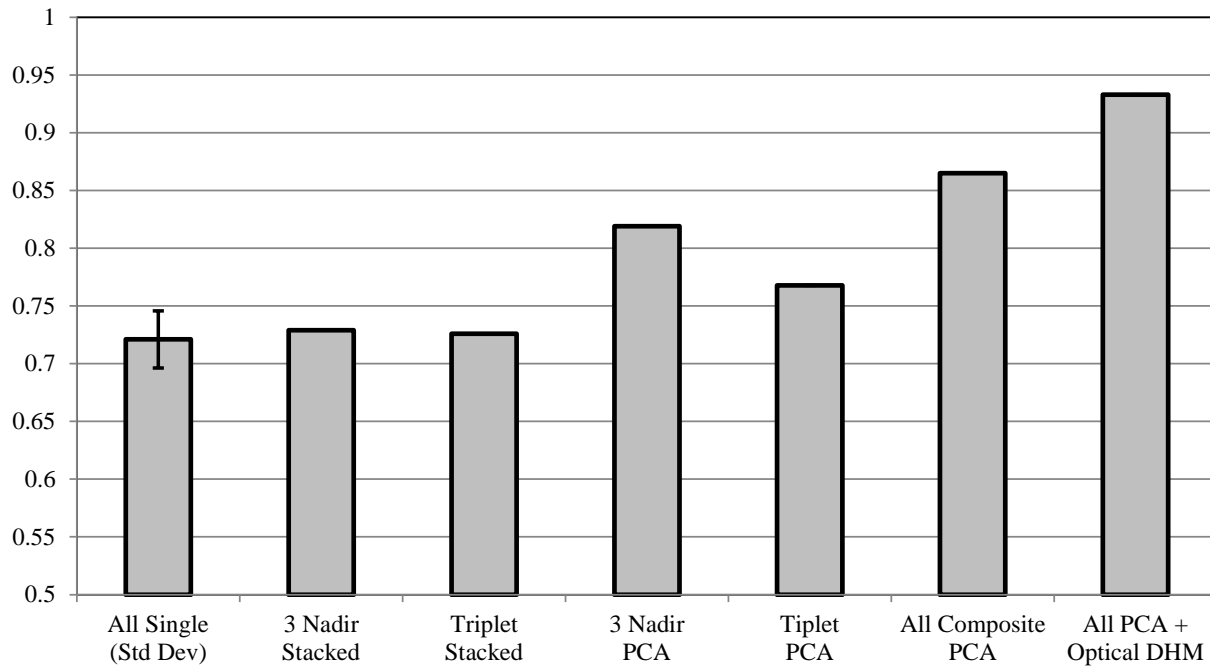
Figure 6.17: Kappa coefficient of all composite multi-angle classification experiments.

In the first three experiments, testing the classification performance of the three main data sets without feature extraction, there is no meaningful performance improvement. This is to say that without feature extraction, the multi-angle and multi-temporal data sets are only a collection of images without additional information available beyond that of a single image.

The next two experiments, "3 Nadir PCA" (#17) and "Triplet PCA" (#18), both show improved classification accuracy over that of the non-feature extracted data sets. Adding the multi-temporal information (#17) to that of the multispectral WorldView-2 imagery, provides an improvement of 14% over the average classification performance of an individual image. Similarly, the multi-angle information (#18) provides an improvement of 7% over the performance of a single image. This is similar to the performance improvement as seen in the in-track experiments (Section 6.3) using PCA feature extraction.

Experiment #19 uses the information features extracted from the full composite multi-angle data set. This process, in essence, combines both the multi-temporal and multi-angle information, shown in experiments #17 and #18, into a single set of features. By combining these two infor-

mation sources, both shown to improve the classification of the 12 class ground survey described above, an improvement of 20% over the average classification performance of a single image can be achieved.

In the final experiment (#20), an additional feature of height in included in the classification. This provides an additional mechanism to distinguish between spectrally similar, but vertically structured, classes, such as roof tops and asphalt. In this final case, a total improvement of 29% over the base case of a single multispectral WorldView-2 image is achieved.

The classification map from this final experiment is shown in Figure 6.18. This map is provided without any post-processing, such as that application of a minimum mapping unit, or manual cleaning. In general, the image shows good identification of the various classes. Specifically, the asphalt roadway and parking lots are generally delineated despite the shared spectral (asphalt) and temporal (moving vehicles) characteristics. In addition, the differentiation between pitched (generally residential), flat (generally commercial), and the skyscrapper classes is evident throughout the residential area and also in the commercial area to the top right of the imager where the building complex has a mixed roofline (pitched and flat). During the course of the composite multi-angle, there was also significant construction present in the target area. This is evident in the bottom right of the image where a pitched roof structure (the roof is a sequence of half domes) has significant noise due to the addition of solar panels part way through the composite multi-angle sequence. A more detailed analysis of various subsections will be discussed below.

To further break down the class-by-class performance, a measurement of classification performance for each of the classes is provided, as measured by F1 score, in Figure 6.19. This shows several features of interest. First, as was the case in the in-track analysis (and as dictated by the information available for classification), bridge classification requires the height feature while the skyscraper feature is significantly improved through the use of this additional information. In addition, both of the tree classes show significant improvement with the addition of the composite multi-angle feature set and also show some improvement (though less than may be expected) with the addition of height information.
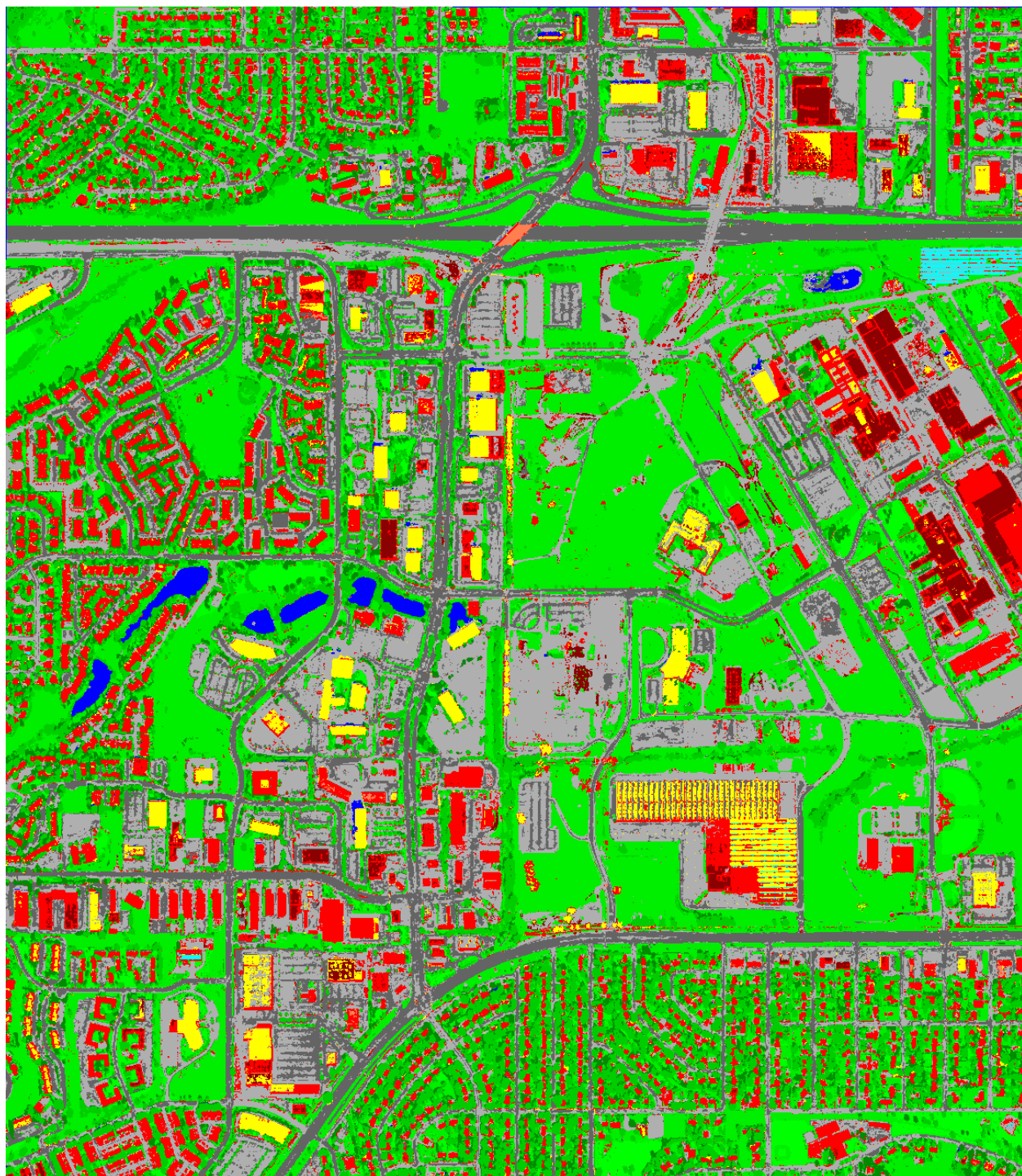
Figure 6.18: Classification image of the Lakewood, CO scene. Color legend: ■ = Pitched Roof, ■ = Flat Roof, ■ = Asphalt Roadway, ■ = Parking Lot, ■ = Bridge, ■ = Solar Panels, ■ = Skyscrapers, ■ = Water, ■ = Conifer, ■ = Deciduous, ■ = Natural Grass, ■ = Maintained Grass.
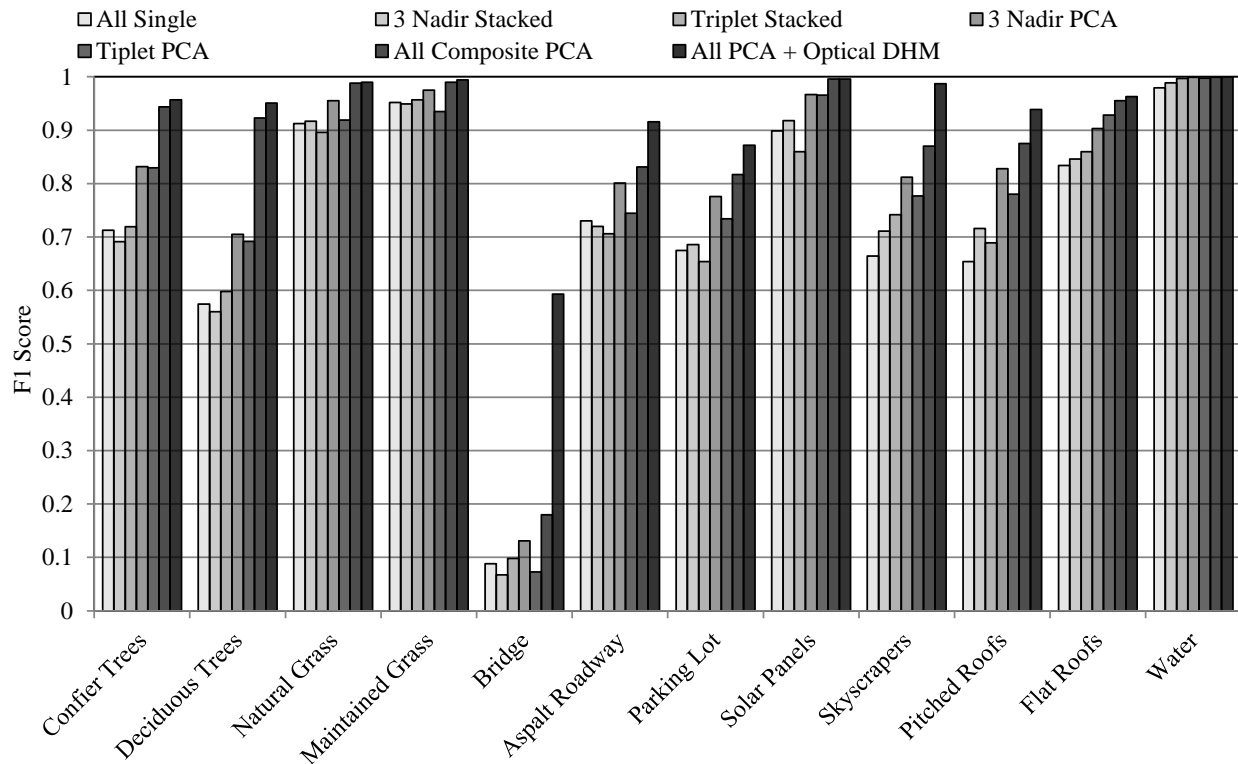
Figure 6.19: F1-score of the individual classes used in the composite multi-angle experiments.

Each of these data sets was designed to segment one type of information from the full complexity of the composite multi-angle data set. This information is then extracted by the PCA. As shown in the previous section, this can also be accomplished by physically modeling the multi-angle structure of the data set. However, the statistical nature of the PCA processing allows the straight-forward development of multiple data sets in order to fill out the parametric study.

This additional information can also provide a more subtle impact on issues that must be dealt with, or are often ignored, in single-image classification. For example, vehicles are often treated as an individual class when analyzing a single image. This is necessary because there is no other information to indicate the land-cover or land-use classification of the area occupied by the vehicle. In some cases, the vehicle is the information of interest, such is the case when trying to determine the number of vehicles in a parking area. However, this use case would more appropriately be treated by object detection. Also, shadows are often treated as a separate class. In a similar way,

this is not the land-use or land-cover information that is of interest. The information of interest is the material on the ground inside of the shadow.

By using the full information available in a composite multi-angle data sequence, these effects can be addressed to some degree. In the composite multi-angle imagery, both of the cases mentioned become transient in that both vehicles and shadows move image to image. In the case of vehicles, this is simply due to the real movement on the surface between images. For shadows, this is an effect of both changing solar illumination as well as observation angle. When observing from the forward scattering direction (opposite the sun), shadows are typically longer and the shaded side of objects are observed. The opposite is true when observing from the backscatter direction. This, combined with the changes in solar angle throughout the year, produce significantly varying shadows between the images of the composite multi-angle sequence.

The following examples illustrate the use of the composite multi-angle sequence in order to leverage these transient effects for improved LU/LC classification. In each of the cases, the resulting map is more representative of the actual land-cover or land-use than is available through the use of a single image.

Figure 6.20 (displayed over a two consecutive pages) shows a residential area from the bottom part of the full target area. The two page figure consists of an RGB aerial image (6.20a), a spectral classification map based only on data from the eight multispectral WorldView-2 bands (6.20b), two classification images (6.20c, and 6.20d)from the final classification experiment (#20) using all the available information from the composite multi-angle data.

As can be seen in the RGB image, the area is mainly composed of residential housing with a combination of conifer and deciduous trees as well as a few parking lots on the top and bottom of the region. In the spectral-only classification image, there is significant noise between the two types of trees (■ = Conifer and ■ = Deciduous), the various building types (■ = Pitched Roof, ■ = Flat Roof, and ▢ = Skyscrapers), as well as the asphalt surfaces (■ = Asphalt Roadway, ▢ = Parking Lot). However, when you include the composite multi-angle information, the tree, building, and asphalt classes become much more clearly identified. In the case of the two asphalt

classes, Figure 6.20c identified much of the areas near the side of the residential roadways as parking areas due to the intermittent presence of vehicles throughout the composite sequence. However, the actual definition and validation of this type of classification is difficult. Therefore, in Figure 6.20d, the classes are merged to more accurately show the pixel level classification accuracy of the full scheme. In this form, the residential building are accurately delineated from the surrounding tree cover. In addition, the flat roof commercial buildings are accurately delineated from the pitched roof structures. The only tall structure in the image, the A-frame shaped church building in the lower-left of the image, is accurately extracted as a skyscraper.

In the next example, shown in Figure 6.21, a mixed-use area from the left side of the target area is shown. It consists of open space, commercial area, and high density residential buildings. In Figure 6.21a, a classification map using only the eight band multispectral information is shown. In this case there is significant confusion between the concrete parking areas of the high-density residential buildings and the concrete-topped skyscrapers (▭). In addition, many of the high-density residential buildings are misclassified as parking lots (▭). Figure 6.21b shows the resolution between these spectrally overlapping classes that is created when using the composite multi-angle information. There is still some confusion in the image around the edge of the high-density residential buildings, but these buildings are accurately classified as pitched roof structures. In addition, the concrete parking area (with covered parking areas) to the top of the image are now accurately identified. In addition, the parking areas are accurately delineated from the trafficked paved areas (including both roadways and the areas between parking stalls in parking lots). Also of note, the building in the far bottom right of the classification image is accurately depicted in Figure 6.21b.

Figure 6.22 illustrates shadow mitigation for three large buildings. The RGB satellite (2 m) images of each building are shown on the left (Figures 6.22a, 6.22c, and 6.22e). The corresponding classification images, extracted from the classification image of experiment (#20) are shown on the right (Figures 6.22b, 6.22d, and 6.22f). In each of the three image sets, the classification image provides a much clearer picture of the land-cover present in the shaded portion of the RGB images. In the top image (Figures 6.22a and 6.22b), the small lake to the top of the building is shaded by

(a)

(b)

(c)
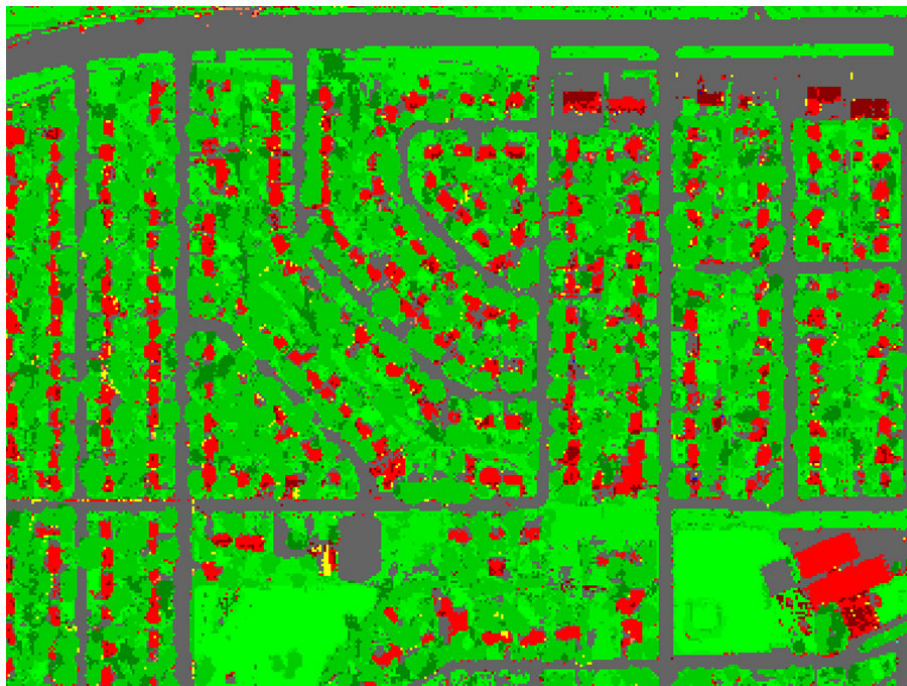
(d)

Figure 6.20: Classification example of a residential area with an aerial image (a) for reference. The three classification examples shown use only the spectral information ((b)) and the full PCA with DHM information stack ((c) and (d)). Image (d) shows the same map as in image (c) but with the two roadway classes merged. (applicable color legend is in Figure 6.18).

(a)

(b)

Figure 6.21: Classification example of a mixed-use area showing the significant improvement resolving classes when the multi-angle information is included. Classification image (a) used the eight band multispectral information which the improved image (b) used the full PCA and DHM image stack. (applicable color legend is in Figure 6.18).

the building. The classification image accurately depicts the shaded area as a combination of water and vegetation. Similarly, the middle image (Figures 6.22c and 6.22d), show the shaded parking area correctly classified as such. The final image set (Figures 6.22e and 6.22f), shows the same effect with the shaded area classified as both parking area and trees. The diagonal delimitation between the parking and building classifications is also generally correct, with one of the main differences between the classes being the height above the surface. In this case, the ground level gradually falls from the left side of the image to the right, so the change in classification between surface pavement and raised building is generally correct.

By leveraging the multi-angle information present in a composite multi-angle WorldView-2 sequence, an improvement of up to 30% in land-cover classification is available and provides the additional capability of mitigating transient effects to more accurately delineate LU/LC classes. In the examples shown above, it was possible to differentiate between parking areas and trafficked paved areas. This includes the differentiation between areas of a parking lot where vehicles travel as compared to where they park. Additionally, the mitigation of shadows of an image can be addressed through the use of the composite multi-angle sequence data allowing for a more accurate LU/LC classification separate from the shadowing effects on the ground.

(a)

(b)

(c)

(d)

(e)

(f)
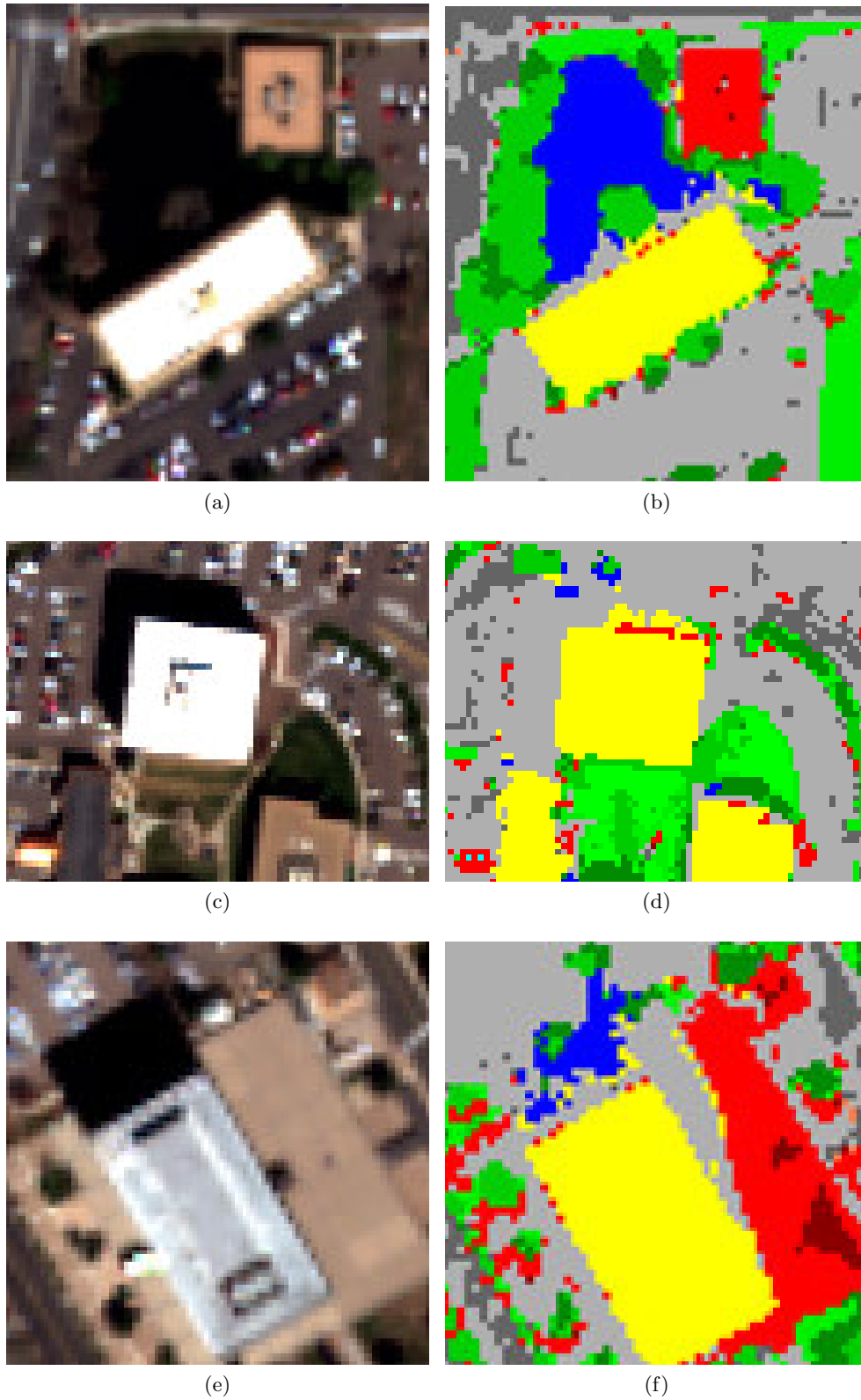
Figure 6.22: Classification example of view-angle and shadow suppression (applicable color legend is in Figure 6.18).

Chapter 7

Conclusions

The studies presented as part of this dissertation cover two major research problems:

(1) Land-use/land-cover classification multispectral model portability.

(2) Multi-angle land-use/land-cover classification improvement.

Each of these problems are uniquely addressed using two types of multi-angle data: in-track and composite. In the former case, the data is composed of multiple shots collected sequentially (within minutes) during a single overflight of the satellite. In the latter case, the data is composed of many individual shots collected over a period of time (1.5 years in the presented data) from various angles.

This research explored methods to leverage the capabilities and characteristics of these data sets to improve current capabilities. Specifically, the two data sets encode multi-angle information that can be extracted and leveraged to improve classification LU/LC classification (problem #2 above) or used as a simplified test case (problem #1 above) in order to develop methods for creating more stable and portable classification models. A brief summary the results from each project is provided below.

7.1 Multispectral Model Portability

The model portability studies presented in this dissertation leverage the unique properties of multi-angle data to segment and measure the improvement in various noise sources complicating

the general use of a Random Forest machine model across many images. Specifically, because the in-track multi-angle data was collected during a single satellite overflight, the collected data space is distorted only by the angular dependent scattering effects of the atmosphere and real BRDF type observed reflectance changes on the surface. Similarly, the composite multi-angle data set, while possessing all of the noise sources typically seen between multiple images, is constrained to a single target area. Therefore, methods to improve model portability can be realistically tested against the more complex data set.

Testing of five data sets, including digital number (DN), top-of-atmosphere reflectance (TOA), atmospheric compensation (AC), DN band ratios, and AC band ratios, on the simplified test case of in-track multi-angle data showed that compensation for both atmospheric scattering and illumination variation is necessary to maximize the portability of a multispectral Random Forest model. Additionally, the model portability testing preferentially favored training in the solar backscatter region. When then effects are maximized, the portability of a multispectral model improved from no portability (completely losing all classification accuracy across the in-track sequence) to a reduction of 10% in kappa coefficient.

When the same analysis was applied to the more complex composite multi-angle data set, similar results were observed. However, the added noise of seasonality limited the portability performance to an overall reduction in classification accuracy across the composite multi-angle data sequence of approximately 20% in the best cases. Additionally, it was shown that, at the most basic level, the correction of DN data to TOA reflectance is a minimal first step to multi-image analysis.

In the future, additional work should be directed at addressing data space distortions due to seasonality and BRDF in order to compensate for the remaining real distortions in the image data space. A possible avenue for this is the statistical domain adaptation techniques described earlier. However, this will likely have to be a class level correction in order to accurately address the differences in seasonality changes and BRDF between classes (i.e. asphalt versus grass). Alternatively,

the use of non-spectral features (such as texture) could improve model portability by removing the spectral dependence of the model all together.

7.2   LU/LC Classification Improvement

These studies showed that there is significant improvement in classification accuracy available from leveraging the spatial and spectral data in both in-track multi-angle and composite multi-angle WorldView-2 image sequences. Specifically, the in-track spatial and spectral information provide 27% and 14% improvement, respectively, in kappa coefficient over the base case of a single nadir multispectral image while the composite multi-angle information collectively improved classification performance 29% over the base case of a classification using the eight bands of a single WorldView-2 multispectral image. In addition to providing the foundation for feature extraction methods to improve classification, the features extracted from the multi-angle data provides the capability to differentiate classes that are generally difficult to distinguish, such as skyscrapers, bridges, high volume highways, parked cars, trafficked asphalt, and parking lots.

Eight spatial classification experiments were presented using a combination of spectral, textural, morphological, and height information. Within these data sets, the DHM provided the largest benefit to the overall classification. It was particularly useful for several classes, including skyscrapers, building, residential homes, and trees. Healthy vegetation relied mainly on the multispectral information and very little on the derived data sources. Although texture was the least significant data set to the overall classification experiment as measured by feature contribution, it did play a uniquely strong role in vehicle detection where the addition of the texture data set alone increased class F-measure from 25% to 81%.

Five spectral classification experiments were separately presented using a nadir multispectral image, the full multi-angle multispectral data set, and three feature extraction techniques. These experiments showed that the multi-angle spectral information generally provided significantly improved classification performance. While the spectral information provided the strongest feature contribution to the overall classification, specific classes benefited from the unique aspects of the

multi-angle information. The derived y-fit error data provided a uniquely strong signal for some of the classes in the spectral experiments. The parked car class does not have a standard spectral signature and, therefore, had very poor class F-measure using only the nadir spectral information (24%). The final class F-measure was increased by 79% with the inclusion of the y-fit error information describing the solar reflection variability across view-angle. The multi-angle spectral variability also provided strong feature contribution for the classes with high temporal variability, such as the roadway classes.

Seven composite multi-angle classification experiments were presented using a parametric approach to measure the impact of the spectral, BRDF type multi-angle, multi-temporal, and height information. The experiments showed that, when feature extraction techniques are appropriately applied, classification performance can be improved 7% through the inclusion of multi-angle information and 14% through the inclusion of multi-temporal information. When these data sources are combined with an optically extracted height model, a total improvement of 29% was observed over the base case of a single multispectral WorldView-2 image. In the presented case, this provided an overall classification performance of 0.93 kappa across 12 classes. This research also showed the promise of this type of LU/LC classification to suppress noise sources, such as building lean and transient vehicles, to accurately classify surfaces such as parking lots.

In the future, further benefit may be realized by physically modeling the spatial, spectral, and temporal information presented in the composite classification studies. This type of physically representative information may prove particularly powerful for identifying land use classes, such as highways, bridges, residential roadways, and parking lots, where the non-spectral information is necessary to accurately identify the classes. Additionally, expanding the multi-angle analysis beyond the 30°off-nadir limit used in this study could benefit classes with significant BRDF. If high off-nadir images can be accurately registered with the near nadir portion of the multi-angle sequence, more of the BRDF information would be represented and, theoretically, increase retrieval accuracy for classes with a BRDF signal.

Bibliography

[1] Digitalglobe base product series faq. Technical report, DigitalGlobe, 2012.

[2] Worldview-3 design and specifications. Technical report, DigitalGlobe, Longmont, CO, 2012.

[3] A Abuelgasim. Classification of asas multiangle and multispectral measurements using artificial neural networks. Remote Sensing of Environment, 57(2):79–87, August 1996.

[4] American Institute of Biological Sciences. Land use and habitat alteration: Report from a neon science workshop. Technical report, AIBS, Washington, DC, 2004.

[5] D. Anding and R. Kauth. Estimation of sea surface temperature from space. Remote Sensing of Environment, 1(4):217–220, December 1970.

[6] A Antonarakis, K Richards, and J Brasington. Object-based land cover classification using airborne lidar. Remote Sensing of Environment, 112(6):2988–2998, June 2008.

[7] L.E. Band. Topographic partition of watersheds with digital elevation models. Water resources research, 22(1):15–24, 1986.

[8] M. J. Barnsley, D. Allison, and P. Lewis. On the information content of multiple view angle (mva) images. International Journal of Remote Sensing, 18(9):1937–1960, June 1997.

[9] M. J. Barnsley, J. J. Settle, M. A. Cutter, D. R. Lobb, and F. Teston. The proba/chris mission: a low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. IEEE Transactions on Geoscience and Remote Sensing, 42(7):1512–1520, July 2004.

[10] R. Bellman. Adaptive Control Processes: A Guided Tour. Princeton University Press, 1961.

[11] Moshe Ben-Ezra, Jiaping Wang, and Bennett Wilburn. An led-only brdf measurement device. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, number i, pages 1–8. IEEE, June 2008.

[12] J.A. Benediktsson, Jocelyn Chanussot, and Mathieu Fauvel. Multiple classifier systems in remote sensing : From basics to recent developments. In Multiple Classifier Systems, pages 501–512, Prague, Czech Republic, 2007. Springer.

[13] Alexander Berk. Modtran4 radiative transfer modeling for atmospheric correction. In Proceedings of SPIE, volume 3756, pages 348–353. SPIE, 1999.

[14] L.S. Bernstein, S.M. Adler-Golden, R.L. Sundberg, R.Y. Levine, T.C. Perkins, A. Berk, A.J. Ratkowski, G.W. Felde, and M.L. Hoke. A new method for atmospheric correction and aerosol optical property retrieval for vis-swir multi- and hyperspectral imaging sensors: Quac (quick atmospheric correction). In 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005., volume 5, pages 3549–3552. IEEE, 2005.

[15] C. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, New York, 1996.

[16] G.W. Bothwell, E.G. Hansen, R.E. Vargo, and K.C. Miller. The multi-angle imaging spectroradiometer science data system, its products, tools, and performance. IEEE Transactions on Geoscience and Remote Sensing, 40(7):1467–1476, July 2002.

[17] Leo Breiman. Random forests. Machine Learning, 45(1):5–32, 2001.

[18] Leo Breiman and Adele Cutler. Random forests.

[19] G. Camps-Valls and Lorenzo Bruzzone. Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 43(6):1351–1362, June 2005.

[20] Jing M. Chen. Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. Remote Sensing of Environment, 84(4):516–525, April 2003.

[21] Chris Comp. Radiometric use of worldview-2 imagery technical note. Technical report, DigitalGlobe, 2010.

[22] Russell G Congalton and Kass Green. Assessing the Accuracy of Remotely Sensed Data Principles and Practices. CRC Press, 2009.

[23] T. Cooley, G.P. Anderson, G.W. Felde, M.L. Hoke, A.J. Ratkowski, J.H. Chetwynd, J.A. Gardner, S.M. Adler-Golden, M.W. Matthew, A. Berk, L.S. Bernstein, P.K. Acharya, D. Miller, and P. Lewis. Flaash, a modtran4-based atmospheric correction algorithm, its application and validation. In IEEE International Geoscience and Remote Sensing Symposium, volume 3, pages 1414–1418. IEEE, 2002.

[24] J. Crespo, Jean Serra, and Ronald W. Schafer. Theoretical aspects of morphological filters by reconstruction. Signal Processing, 47(2):201–225, November 1995.

[25] Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle. Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas. IEEE Transactions on Geoscience and Remote Sensing, 46(5):1416–1427, May 2008.

[26] Pablo DAngelo and Georg Kuschk. Dense multi-view stereo from satellite imagery. In IEEE International Geoscience and Remote Sensing Symposium, pages 6944–6947, 2012.

[27] M. Davidson and P. Vuilleumier. Note on chris acquisition procedure and image geometry. Technical report, ESA, 2005.

[28] Ruth DeFries, Matthew C. Hansen, and John Townshend. Global discrimination of land cover types from metrics derived from avhrr pathfinder data. Remote Sensing of Environment, 54(3):209–222, December 1995.

[29] Fabio Del Frate, Riccardo Duca, Pasquale Sellitto, and Domenico Solimini. Use of chris proba images for land use products. 2007 IEEE International Geoscience and Remote Sensing Symposium, pages 2804–2807, 2007.

[30] Fabio Del Frate, Fabio Pacifici, Giovanni Schiavon, and Chiara Solimini. Use of neural networks for automatic classification from high-resolution images. IEEE Transactions on Geoscience and Remote Sensing, 45(4):800–809, April 2007.

[31] Kurt DeVenecia, Stewart Walker, and Bingcai Zhang. New approaches to generating and processing high resolution elevation data with imagery. Photogrammetric Week 07, pages 297–308, 2007.

[32] D.J. Diner, J.C. Beckert, T.H. Reilly, C.J. Bruegge, J.E. Conel, Ralph A. Kahn, J.V. Martonchik, T.P. Ackerman, R. Davies, S.A.W. Gerstl, H.R. Gordon, Jan-Peter Muller, R.B. Myneni, P.J. Sellers, B. Pinty, and M.M. Verstraete. Multi-angle imaging spectroradiometer (misr) instrument description and experiment overview. IEEE Transactions on Geoscience and Remote Sensing, 36(4):1072–1087, July 1998.

[33] Riccardo Duca and Fabio Del Frate. Hyperspectral and multiangle chris-proba images for the generation of land cover maps. IEEE Transactions on Geoscience and Remote Sensing, 46(10):2857–2866, October 2008.

[34] Giles M. Foody. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote, 70(May):627–633, 2004.

[35] Mark A. Friedl, D.K McIver, J.C.F Hodges, X.Y Zhang, D. Muchoney, A.H. Strahler, Curtis Woodcock, Sucharita Gopal, Annemarie Schneider, A Cooper, A Baccini, F Gao, and C. L. Barker Schaaf. Global land cover mapping from modis: algorithms and early results. Remote Sensing of Environment, 83(1-2):287–302, November 2002.

[36] G.T. Georgiev, C.K. Gatebe, J.J. Butler, and M.D. King. Brdf analysis of savanna vegetation and salt-pan samples. IEEE Transactions on Geoscience and Remote Sensing, 47(8):2546–2556, August 2009.

[37] P Gislason, J.A. Benediktsson, and J Sveinsson. Random forests for land cover classification. Pattern Recognition Letters, 27(4):294–300, March 2006.

[38] I. F. Grant. Investigation of the variability of the directional reflectance of Australian land cover types. Remote Sensing Reviews, 19(1-4):243–258, December 2000.

[39] Norbert Haala and Claus Brenner. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing, 54(2-3):130–137, July 1999.

[40] Li Haitao, Gu Haiyan, Han Yanshun, and Yang Jinghui. Fusion of high-resolution aerial imagery and lidar data for object-oriented urban land-cover classification based on svm. In ISPRS Workshop on Updating Geo-spatial Databases with Imagery, number 1, Urumchi, China, 2007. Citeseer.

[41] Robert M. Haralick. Statistical and structural approaches to texture. Proceedings of the IEEE, 67(5):786–804, 1979.

[42] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 3(6):610–621, November 1973.

[43] Otto P Hasekamp and Jochen Landgraf. Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements. Applied Optics, 46(16):3332, 2007.

[44] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: data mining, inference, and prediction. Springer, 2nd edition, 2009.

[45] Eugene Hecht. Optics. Addison Wesley Publishing Company, 3rd edition, 1997.

[46] Wei Hu. Identifying predictive markers of chemosensitivity of breast cancer with random forests. Journal of Biomedical Science and Engineering, 3(1):59–64, 2010.

[47] K. Jacobsen. Characteristics of very high resolution optical satellites for topographic mapping. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVIII-4/:137–142, September 2012.

[48] SK Jenson and JO Domingue. Extracting topographic structure from digital elevation data for geographic information system analysis. Photogrammetric Engineering and remote sensing, 54(11):1593–1600, 1988.

[49] Jing Jiang. Domain Adaptation in Natural Language Processing. PhD thesis, University of Illinois at Urbana-Champaign, 2008.

[50] Goo Jun and Joydeep Ghosh. Spatially adaptive classification of land cover with remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 49(7):2662–2673, July 2011.

[51] Ralph A. Kahn, Pranab Banerjee, and Duncan McDonald. Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. Journal of Geophysical Research, 106(D16):18219–18238, 2001.

[52] Miles V. Klein and Thomas E. Furtak. Optics. John Wiley & Sons, 2nd edition, 1986.

[53] Alexander A. Kokhanovsky and Gerrit de Leeuw. Satellite Aerosol Remote Sensing Over Land. Springer, 2009.

[54] Tatjana Koukal and Clement Atzberger. Potential of multi-angular data derived from a digital aerial frame camera for forest classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1):30–43, February 2012.

[55] Keith Krause. Radiometric use of quickbird imagery. Technical Report July, DigitalGlobe, Longmont, CO, 2005.

[56] Bor-Chen Kuo. Improved Statistics Estimation and Feature Extraction for Hyperspectral Data Classification. PhD thesis, Purdue University, 2001.

[57] David A Landgrebe. Signal Theory Methods in Multispectral Remote Sensing. John Wiley & Sons, Hoboken, 2003.

[58] Craig A Lee, Samuel D Gasster, Antonio Plaza, Chein-I Chang, and Bormin Huang. Recent developments in high performance computing for remote sensing: A review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4(3):508–527, September 2011.

[59] Hae-Yeoun Lee, Taejung Kim, Wonkyu Park, and Heung Kyu Lee. Extraction of digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry. Image and Vision Computing, 21(9):789–796, September 2003.

[60] Krista R Lee. Using multi-angle worldview-2 imagery to determine ocean depth near oahu, hawaii. Master's thesis, Naval Postgraduate School, 2012.

[61] Aude Lemonsu, Alexandre Leroux, Stephane Belair, Serge Trudel, and Jocelyn Mailhot. A general methodology of urban cover classification for atmospheric modelling. In Sixth Symposium on the Urban Environment, number July, 2006.

[62] Shunlin Liang, Senior Member, Hongliang Fang, and Mingzhen Chen. Atmospheric Correction of Landsat ETM + Land Surface Imagery Part I : Methods. IEEE Transactions on Geoscience and Remote Sensing, 39(11):2490–2498, 2001.

[63] Andy Liaw. Package "randomforest". Technical report, 2011.

[64] V Liesenberg, L Galvao, and F Ponzoni. Variations in reflectance with seasonality and viewing geometry: Implications for classification of brazilian savanna physiognomies with misr/terra data. Remote Sensing of Environment, 107(1-2):276–286, March 2007.

[65] Yang Liu. Validation of multiangle imaging spectroradiometer (misr) aerosol optical thickness measurements using aerosol robotic network (aeronet) observations over the contiguous united states. Journal of Geophysical Research, 109(D6):1–9, 2004.

[66] Nathan Longbotham, Chuck Chaapel, Laurence Bleiler, Chris Padwick, William J. Emery, and Fabio Pacifici. Very high resolution multiangle urban classification analysis. IEEE Transactions on Geoscience and Remote Sensing, 50(4):1155–1170, April 2012.

[67] Nathan Longbotham, Fabio Pacifici, Taylor Glenn, Alina Zare, Michele Volpi, Devis Tuia, Emmanuel Christophe, Julien Michel, Jordi Inglada, Jocelyn Chanussot, and Qian Du. Multi-Modal Change Detection, Application to the Detection of Flooded Areas: Outcome of the 2009/2010 Data Fusion Contest. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1):331–342, February 2012.

[68] Anjum Mahtab, V. N. Sridhar, and Ranganath R. Navalgund. Impact of surface anisotropy on classification accuracy of selected vegetation classes: An evaluation using multidate multiangular misr data over parts of madhya pradesh, india. IEEE Transactions on Geoscience and Remote Sensing, 46(1):250–258, January 2008.

[69] Giovanni Marchisio, Fabio Pacifici, and Chris Padwick. On the relative predictive value of the new spectral bands in the worldwiew-2 sensor. In 2010 IEEE International Geoscience and Remote Sensing Symposium, pages 2723–2726. IEEE, July 2010.

[70] Stuart E Marsh and Ronald J.P. Lyon. Quantitative relationships of near-surface spectra to landsat radiometric data. Remote Sensing of Environment, 10(4):241–261, December 1980.

[71] J.V. Martonchik, D.J. Diner, Ralph A. Kahn, T.P. Ackerman, M.M. Verstraete, B. Pinty, and H.R. Gordon. Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging. IEEE Transactions on Geoscience and Remote Sensing, 36(4):1212–1227, July 1998.

[72] J. F. Mas and J. J. Flores. The application of artificial neural networks to the analysis of remotely sensed data. International Journal of Remote Sensing, 29(3):617–663, February 2008.

[73] Paul Mather. Computer Processing of Remotely-Sensed Images: An Introduction. John Wiley & Sons, West Sussex, England, 3rd edition, 2004.

[74] Paul M. Mather and Magaly Koch. Computer Processing of Remotely-Sensed Images - An Introduction. John Wiley & Sons, 4th edition, 2011.

[75] Helmut Mayer. Automatic Object Extraction from Aerial ImageryA Survey Focusing on Buildings. Computer Vision and Image Understanding, 74(2):138–149, May 1999.

[76] Ryan A. Mercovich, James Albano, and David Messinger. Techniques for the graph representation of spectral imagery. In 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pages 1–4. IEEE, June 2011.

[77] I.D. Moore, P.E. Gessler, G.A. Nielsen, and G.A. Peterson. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal, 57(2):443–443, 1993.

[78] Emma-claire Mullally, Diarmuid P O Donoghue, Amy J Bohan, and Mark T Keane. Geometric proportional analogies in topographic maps: Theory and application. In Ann Macintosh, Richard Ellis, and Tony Allen, editors, Applications and Innovations in Intelligent Systems XIII, pages 95–108, London, 2006. Springer London.

[79] K. Navulur. Multispectral image analysis using the object-oriented paradigm. CRC, 2007.

[80] Fabio Pacifici, Marco Chini, and William J. Emery. A neural network approach using multiscale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sensing of Environment, 113(6):1276–1292, June 2009.

[81] Chris Padwick, M. Deskevich, Fabio Pacifici, and S. Smallwood. Worldview-2 pan-sharpening. In ASPRS, San Diego, CA, 2010.

[82] Chris Padwick, Justin Luitjens, and Kevin Berce. High performance cloud computing for image processing. In MAPPS/ASPRS 2012 Specialty Conference, Tampa, Florida, 2012.

[83] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, October 2010.

[84] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition, 44(3):678–693, March 2011.

[85] Antonio Plaza, Qian Du, Yang-lang Chang, and Roger L King. High performance computing for hyperspectral remote sensing. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4(3):528–544, September 2011.

[86] Thomas Pock, Lukas Zebedin, and Horst Bischof. Tgv-fusion. In Cristian Calude, Grzegorz Rozenberg, and Arto Salomaa, editors, Rainbow of Computer Science, volume 6570 of Lecture Notes in Computer Science, pages 245–258. Springer Berlin / Heidelberg, 2011.

[87] Jeffrey L. Privette, Thomas F. Eck, and Donald W. Deering. Estimating spectral albedo and nadir reflectance through inversion of simple brdf models with avhrr/modis-like data. Journal of Geophysical Research, 102(D24):29529–29542, 1997.

[88] Jeffrey Lawrence Privette. An efficient strategy for the inversion of bidirectional reflectance models with satellite remote sensing data. PhD thesis, University of Colorado, 1994.

[89] E. Puttonen, J. Suomalainen, T. Hakala, and J. Peltoniemi. Measurement of reflectance properties of asphalt surfaces and their usability as reference targets for aerial photos. IEEE Transactions on Geoscience and Remote Sensing, 47(7):2330–2339, July 2009.

[90] J.R. Quinlan. Induction of decision trees. Machine Learning, 1(1):81–106, October 1986.

[91] D. Riaño, Emilio Chuvieco, Javier Salas, and Inmaculada Aguado. Assessment of different topographic corrections in landsat-tm data for mapping vegetation types. IEEE Transactions on Geoscience and Remote Sensing, 41(5):1056–1061, May 2003.

[92] H C Schau. Statistical filter for image feature extraction. Applied Optics, 19(13):2182, July 1980.

[93] M. Schlerf, W. Verhoef, H. Buddenbaum, J. Hill, Clement Atzberger, and A. Skidmore. Comparing three canopy reflectance models with hyperspectral multi-angular satellite data. In Proceedings of the 10th International Symposium on Physical Measurements and Spectral Signatures in Remote Sensing, volume 1, pages 12–14, Davos (Switzerland), 2007.

[94] JR Schott. Remote sensing: the image chain approach. Oxford University Press, 2007.

[95] Robert A. Schowengerdt. Remote sensing: Models and Methods for Image Processing. Elsevier, 3rd edition, December 2007.

[96] Christopher Small. High spatial resolution spectral mixture analysis of urban reflectance. Remote Sensing of Environment, 88(1-2):170–186, November 2003.

[97] Jeong-Heon Song, Soo-Hee Han, Kiyun Yu, and Yong-Il Kim. Assessing the possibility of land-cover classification using lidar intensity data. In Photogrammetric Computer Vision ISPRS Commission III, volume 34, Graz, Austria, 2002.

[98] Helene Sportouche, Florence Tupin, and Leonard Denise. Building extraction and 3d reconstruction in urban areas from high-resolution optical and sar imagery. In 2009 Joint Urban Remote Sensing Event, pages 1–11. IEEE, May 2009.

[99] A.H. Strahler, JP Muller, W. Lucht, CB Schaaf, T. Tsang, F. Gao, X. Li, P. Lewis, and M.J. Barnsley. Modis brdf/albedo product: algorithm theoretical basis document. Technical Report April 1999, MODIS Science Team, 1999.

[100] L Su, Mark Chopping, Albert Rango, J.V. Martonchik, and D Peters. Support vector machines for recognition of semi-arid vegetation types using misr multi-angle imagery. Remote Sensing of Environment, 107(1-2):299–311, March 2007.

[101] Céline Tison, Florence Tupin, and Henri Maître. A fusion scheme for joint retrieval of urban height map and classification from high-resolution interferometric sar images. IEEE Transactions on Geoscience and Remote Sensing, 45(2):496–505, 2007.

[102] T. Toutin. Elevation modelling from satellite visible and infrared (vir) data. International Journal of Remote Sensing, 22(6):1097–1125, January 2001.

[103] Devis Tuia, Jordi Muñoz Marí, Luis Gómez-Chova, and Jesus Malo. Graph matching for adaptation in remote sensing. IEEE Transactions on Geoscience and Remote Sensing, pages 1–13, 2012.

[104] Devis Tuia, Fabio Pacifici, Mikhail Kanevski, and William J. Emery. Classification of very high spatial resolution imagery using mathematical morphology. IEEE Transactions on Geoscience and Remote Sensing, 47(11):3866–3879, 2009.

[105] Devis Tuia, E. Pasolli, and William J. Emery. Using active learning to adapt remote sensing image classifiers. Remote Sensing of Environment, 115(9):2232–2242, September 2011.

[106] C. J. van Rijsbergen. Information Retrieval. Butterworth, London; Boston, 2nd edition, November 1979.

[107] W. Verhoef. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. Remote Sensing of Environment, 87(1):23–41, September 2003.

[108] W Wanner, A.H. Strahler, B Hu, P. Lewis, Jan-Peter Muller, X Li, C. L. Barker Schaaf, and M. J. Barnsley. Global retrieval of bidirectional reflectance and albedo over land from eos modis and misr data: Theory and algorithm. Journal of Geophysical Research, 102(D14):17143, July 1997.

[109] G.G. Wilkinson. Results and implications of a study of fifteen years of satellite image classification experiments. IEEE Transactions on Geoscience and Remote Sensing, 43(3):433–440, March 2005.

[110] Michael A. Wulder, Joanne C. White, Ross F. Nelson, Erik Næsset, Hans Ole Ørka, Nicholas C. Coops, Thomas Hilker, Christopher W. Bater, and Terje Gobakken. Lidar sampling for large-area forest characterization: A review. Remote Sensing of Environment, 121:196–209, June 2012.

[111] Zhengwei Yang, Patrick Willis, and Rick Mueller. Impact of band-ratio enhanced awifs image to crop classification accuracy. In Pecora 17 The Future of Land Imaging. . . Going Operational, Denver, CO, 2008. American Society for Photogrammetry and Remote Sensing.

[112] Hua Yuan, Xiao-Ping Zhang, and Ling Guan. A statistical approach for image feature extraction in the wavelet domain. In CCECE 2003 - Canadian Conference on Electrical and Computer Engineering), volume 2, pages 1159–1162. IEEE, 2003.

[113] Jixian Zhang. Multi-source remote sensing data fusion: status and trends. International Journal of Image and Data Fusion, 1(1):5–24, March 2010.

[114] Weihua Zhang and D.R. Montgomery. Digital elevation model grid size, landscape representation, and hydrologic simulations. Water Resources Research, 30(4):1019–1028, 1994.

Appendix A

Designing a Random Forest

In the experiments presented in this dissertation, additional features derived from the composite multi-angle imagery, such as spectral pixel variability and pixel height, are used in order to provide more useful information to a machine learning system and to test the portability of a basic multispectral classification model. This is measured in each experiment by using changes in kappa coefficient from a base case experiment. In the classification experiments, this is a WorldView-2 eight band multispectral image. In the model portability experiments, the base case is the training location of the classification model.

In this dissertation, all classification models have been created using the Random Forest algorithm trained with supervised data [17]. In each set of experiments, the classification results were calculated using a single set of parameters for the Random Forest in order to minimize the variability of the results due to the classification algorithm. However, due to the random nature of the algorithm, it is non-deterministic in that, given the same set of input data, the algorithm will produce different results in multiple training/validation experiments [12]. Therefore, it is generally necessary to statistically measure the improvement in kappa coefficient from one experiment run to the next.

The specific choices of Random Forest parameters, most prominently forest size and variable depth, can have a significant effect on the performance of the algorithm. Because of this, there are many methods for optimization of the Random Forest algorithm [18]. Additionally, the authors of

programming packages that implement the Random Forest algorithm have designed optimization routines into their recommended workflows [63].

However, each of these optimization techniques are problem specific. In order to directly compare the classification performance of the experiments in this study, a single set of parameters are chosen based on experimental variability results. In order to understand and minimize the Random Forest variability due to setup parameters, experiments are presented here that map the variable depth and forest size parameter space.

In the Figures A.1, A.2, and A.3, the forest size and variable depth of multiple classification models are presented against the number of tied vectors in the forest. The number of tied vectors is simply a count of the number of trees in the forest that produce no clear winner between two or more classes. In this way, the number of tied vectors described the amount of confusion in the Random Forest - with higher confusion representing a less predictable classification result.

Figures A.1 through A.3 provide plots of both the Random Forest size and the Random Forest variable depth parameters as a function of tied vectors in the resulting forest. The variable depth space shows that the number of tied vectors is minimized at approximately a variable depth of 4 (in the presented eight feature data space), with no penalty for deeper trees. Similarly, the plots show that the Random Forest size should be greater than a few hundred trees, with the small benefits attained for the larger computational complexity of exceedingly large forests. However, it is also shown that choosing a very small number of trees for a Random Forest produces a particularly ambiguous machine learning model. This can be particularly detrimental in cases where a single model is used to make accuracy improvement determinations between experiments with relatively close kappa coefficient values.

Figure A.3 shows a surface plot from which the two box plots were taken with the variability in tied vectors averaged in order to facilitate presentation in the surface format. The same trends can be seen across the two parameter data space with a significant increase in the variability of each model when the number of trees is very small and little real dependence on the allowed variable depth.
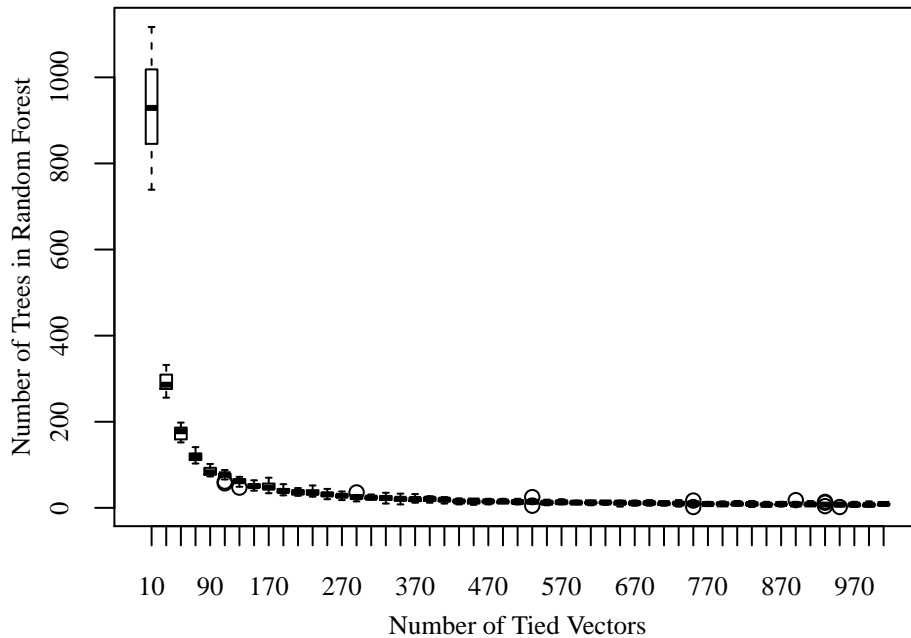
Figure A.1: The number of tied vectors by number of trees in the Random Forest over all variable depth parameters.
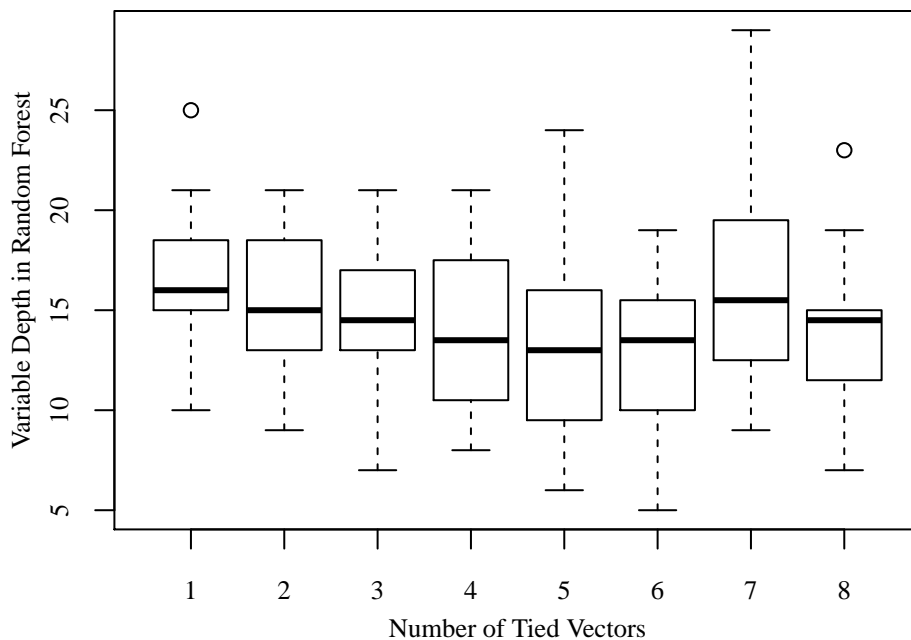

Figure A.2: The number of tied vectors by variable depth (maximum eight WorldView-2 spectral bands) over all forest size parameters.
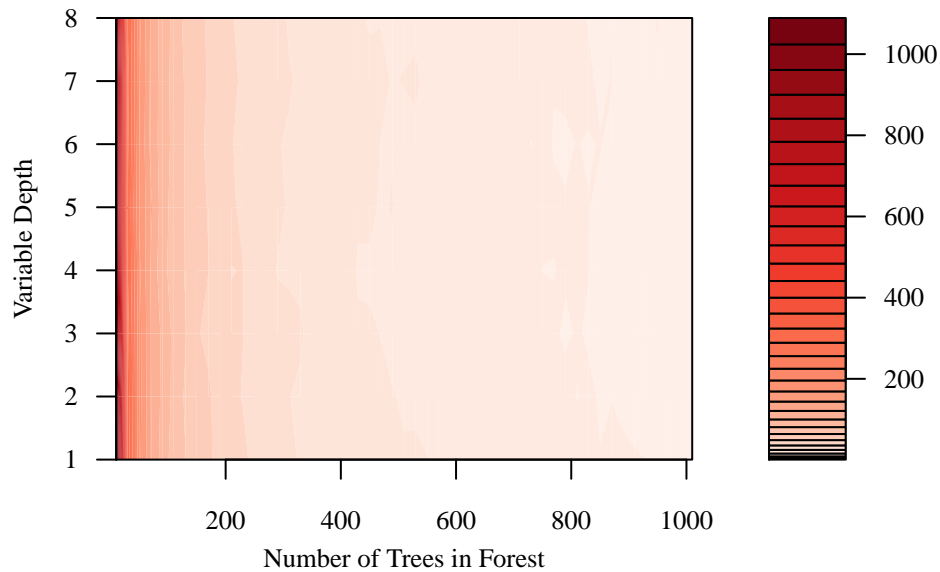
Figure A.3: A surface plot of similar information showing the mean number of tied vectors (over 50 repeated experiments) for the same two Random Forest parameters.

From these plots we can determine some basic guidance for creating reasonable Random Forest model. Based on the above experiments, the number of variables randomly sampled as candidates at each split is $sqrt(n)$, where n is the number of features. Additionally, the number of trees grown as part of each Random Forest is set to 300. This is, in fact, very similar to the default setting of the Random Forest implementation provided by Andy Liaw [63]. However, other implementations, by default, set the forest size to 10 trees. This results in a much faster algorithm, but comes at the cost of additional and significant uncertainty in the reported classification performance.