TOWARD THE DETECTION OF LANDSCAPE FEATURES: CLUSTERING 3D POINTS USING SPATIAL AND THEMATIC CHARACTERISTICS

By

Boleslo Edward Romero

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Geography

ABSTRACT

TOWARD THE DETECTION OF LANDSCAPE FEATURES: CLUSTERING 3D POINTS USING SPATIAL AND THEMATIC CHARACTERISTICS

By

Boleslo Edward Romero

The study of Geography generally concerns phenomena at or near the surface of the earth. High resolutions of 3D quantitative and qualitative data can represent such phenomena as objects or fields. The data can be grouped to reveal representations of contiguous regions of spatial and thematic homogeneity. My thesis is concerned with finding groups of 3D points with similar locations, spatial relationships, and thematic values of spectral reflectance. To accomplish this successfully, I synthesized elements of two geographic theories: point aggregation from cartographic generalization and hierarchical geographic ontology. My experimental design used synthetic 3D point data with spectral values. I employed the multi-dimensional Mean Shift clustering technique from the discipline of Computer Vision, and adapted a 3D range image segmentation accuracy assessment technique. I also contributed new techniques for segmentation quality assessment including two area under the curve indices and the development of new segmentation surface plots. Experimental evaluations included comparisons of the Mean Shift results with K-means clustering results, spatial resolution results, noise evaluation results, and the results of an alternative color configuration. I modified the variable sets to address uneven lighting conditions and employed the experimental methods to grouping real-world terrestrial LiDAR scan data. Though my new spatial relationship variable needs improvement, the methods yielded groups of points representing features in the LiDAR data and provided evidence of the potential for grouping richly attributed 3D points that represent geographic features.

Copyright by

BOLESLO EDWARD ROMERO

DEDICATION

This thesis is dedicated to my mother, my father, my grandparents, and Tammy for their positive influence, steady encouragement, and unconditional love.

ACKNOWLEDGMENTS

Dr. Ashton Shortridge has been an outstanding advisor. I genuinely value his thoughtful review, critiques, and counsel throughout my research. He has been generous with his time, shared many insights, and expressed continued enthusiasm for the practice of scholarship. I also appreciate the contributions of my committee members, Dr. David Lusch and Dr. Bruce Pigozzi. They kindly shared their time and ideas to improve my research. Dr. Tamara Bray offered an opportunity for me to participate in a truly inspiring field research project. Dr. Tim Wawrzyniec, Jed Frechette, and the UNM LiDAR Lab provided detailed guidance and field equipment for the LiDAR case studies. Three other scholars, Dr. Adam Hoover, Dr. Ilan Shimshoni, and Dr. George Stockman, all provided consultation at critical stages. I am grateful for the help and partnership that has been offered by all of the people above and many other friends as well. I express my gratitude to the Michigan State University as a whole and to the Department of Geography faculty and staff for support throughout the course of my thesis.

TABLE OF CONTENTS

LIST OF	TABLES	ix
LIST OF	FIGURES	xi
KEY TO	SYMBOLS OR ABBREVIATIONS	xiii
Chapter 1	I - INTRODUCTION	1
1.1	Major Concepts	1
1.2	Focus of Work	2
1.3	Contributions	4
1.4	General Hypotheses	6
Chapter 2	2 - CONCEPTUAL FRAMEWORK	7
2.1	Overview	7
2.2	Geographic Representation	7
2.3	Cartographic Generalization	10
2.4	Geographic Ontology	11
2.5	Synthesis of Cartographic Generalization and Geographic Ontology	14
2.6	Spatial Characteristics	15
2.6.	1 Derived Spatial Variables - Neighborhood Relationships	15
2.7	Thematic Characteristics	18
2.7.	1 Derived Thematic Variables - Spectral Color Space Alternatives	19
2.8	Classification and Segmentation	19
2.9	Clustering	21
2.9.	1 Hierarchical	22
2.9.2	2 K-means	22
2.9.3	3 ISODATA	24
2.	9.3.1 Mean Shift	25
2.10	Accuracy Assessment	28
2.10	0.1 Types of Clustering Assessment	28
2.10	0.2 Confusion Matrices - Typical Classification Approach	29
2.10	.3 Overlapping Area Matrices - Modified Segmentation Approach	31
2.10	.4 Range Data Segmentation - Extended Metrics	32
2.	10.4.1 Segmentation Performance per Tolerances - Correct Segmentation	34
2.	10.4.2 Instances of Over-segmentation and Under-segmentation	35
2.	10.4.3 Final Classification	36
2.	10.4.4 Segmentation Performance Curves	37
2.	10.4.5 Area Under the Curve Index	38

Chapter 3 - METHODS	
3.1 Methods Overview	
3.2 Data Sources	
3.2.1 Synthetic Data Sets	41
3.2.2 LiDAR Data - Case Studies	
3.3 Processing	
3.3.1 Spatial Variables - Neighborhood Relationshi	ps44
3.3.2 Thematic Variables - Spectral Color Space Al	Iternatives45
3.3.3 Standardized variables	
3.3.4 Variable Sets and List of Experiments	
3.3.4.1 Experiment Set 1 - Mean Shift/K-means	s Clustering Comparison48
3.3.4.2 Experiment Set 2 - Spatial Resolution C	Comparison48
3.3.4.3 Experiment Set 3 - Noise Evaluation	49
3.3.4.4 Experiment Set 4 - Alternative Color Co	onfiguration49
3.3.4.5 Experiment Set 5 - LiDAR Case Studies	s50
3.3.5 Clustering	
3.4 Accuracy Assessment	
3.4.1 Mean Shift Clustering	
3.4.2 Pruned Mode Labeling Modifications	54
3.4.3 Segmentation Performance Classification	
3.4.3.1 Modification to the Reference Column	Metrics55
3.4.3.2 Tolerance Values	
3.4.3.3 Correctly Segmented Metrics	
3.4.3.4 Over-segmented Metrics	
3.4.3.5 Under-segmented Metrics	
3.4.3.6 Final Classification	
3.4.4 Segmentation Performance Curves	60
3.4.5 Area Under the Curve Index	60
3.4.6 Segmentation Surface Plots	61
3.4.7 Statistical Analyses	
3.4.7.1 One-sample t-tests for the Noise Evalua	tion63
3.4.7.2 The ANOVA and Post-ANOVA Tests f	for All Variable Sets63
3.4.7.3 The ANOVA and Post-ANOVA Tests f	for Color Variable Sets64
3.4.7.4 The ANOVA and Post-ANOVA Tests f	for Variable Type64
3.4.8 3D Perspective Plots	
3.4.9 K-means Clustering Comparison	
3.4.10 LiDAR Case Studies	
Chapter 4 - RESULTS and DISCUSSION	
4.1 Overview	
4.2 Synthetic Data Set	
4.3 K-means clustering	
4.4 Spatial Resolution	
4.5 Noise Evaluation	
4.5.1 One-sample t-tests for All Variable Sets at Al	I Levels of Noise
4.5.2 Selection of the 0.02 Noise Level for Addition	nal Analyses87

4.6	Comparison of Multiple Means with Analysis of Variance (ANOVA) tests	88
4.6.	1 The ANOVA and Post-ANOVA Tests for All Variable Sets	
4.6.	2 The ANOVA and Post-ANOVA Tests for Color Variable Sets	92
4.6.	3 The ANOVA and Post-ANOVA Tests for Variable Types	95
4.7	Alternative Color Configuration	97
4.8	LiDAR Case Studies	101
4.8.	1 Indoor Scan	101
4.8.	2 Archaeological Site Scan	107
Chapter	5 - CONCLUSIONS	115
5.1	Overview	115
5.2	Two General Hypotheses	115
5.3	Five Sets of Experiments	115
5.4	Contributions	117
5.5	Contributions to Theory	119
5.6	Future Research	121
REFERE	ENCES CITED	124

LIST OF TABLES

Table 1. Hypothetical overlapping error matrix showing under-segmentation.	32
Table 2. <i>Mr</i> values, proportions of points in each cell relative to the reference totals	33
Table 3. <i>Ms</i> values, proportions of points in each cell relative to the output totals	34
Table 4. Cells with proportions greater than or equal to a tolerance of 0.5	35
Table 5. Cells with proportions greater than or equal to a tolerance of 0.6.	35
Table 6. Color assignment for thematic variables of the initial synthetic data set	42
Table 7. Data sets	44
Table 8. Variable sets, variables and descriptions.	45
Table 9. Combinations of sets of variables, steps, and sequences.	47
Table 10. List of experiments.	47
Table 11. Alternative color configuration values.	50
Table 12. K-means and Mean Shift AUC rankings.	72
Table 13. Low resolution Max. AUC rankings	77
Table 14: High resolution Max. AUC and Avg. AUC rankings	79
Table 15. Comparison of various resolutions' Avg. AUC.	83
Table 16. Mean Avg. AUC and P-values for one-tailed, one-sample t-tests	86
Table 17. ANOVA results for all variable sets.	89
Table 18. ANOVA results for color variable sets.	92
Table 19. Color space value examples favoring clustering of the RGB variable set	94
Table 20. ANOVA results for variable types	95
Table 21. Alternative color Max. AUC and Avg. AUC.	99

Table 22. Indoor LiDAR Max. AUC and Avg. AUC values	
Table 23. Archaeological Site LiDAR Max. AUC and Avg. AUC values	111

LIST OF FIGURES

Figure 1. General summary of theory, processes, methods.	2
Figure 2. Mean vector relationships of 3D points representing an object.	17
Figure 3. Segmented satellite image (true color, false color, and classified)	20
Figure 4. K-means clustering process	23
Figure 5: K-means clustering mismatch between 4 clusters and 3 seeds.	24
Figure 6. Mean Shift on 2D point set, finding modes	26
Figure 7. Points in a 3D LUV color space clustered into 6 groups by the Mean Shift	27
Figure 8. Example segmentation performance curve.	38
Figure 9. Synthetic test cube from the top, front, and left corner	42
Figure 10. Indoor LiDAR data set.	51
Figure 11. Archaeological Site LiDAR data set.	52
Figure 12. Initial noise-free data set.	68
Figure 13. K-means and Mean Shift Segmentation Performance Curves.	70
Figure 14. K-means 3D perspective plots	73
Figure 15. Low resolution segmentation performance curves	77
Figure 16. High resolution segmentation performance curves.	78
Figure 17. High resolution segmentation surface plots.	80
Figure 18. High resolution 3D perspective plots.	81
Figure 19. Examples of adding noise to the spatial data	84
Figure 20. Least significant differences results for all variable sets	90
Figure 21. Least significant differences results for color variable sets	93

Figure 22. Least significant differences results for variable types
Figure 23. Alternative color configuration segmentation performance curves
Figure 24. Alternative color configuration 3D perspective plots
Figure 25. Indoor LiDAR scan, sampled data, and labeled reference groups 102
Figure 26. Indoor LiDAR segmentation performance curve examples
Figure 27. Indoor LiDAR segmentation surface plot examples
Figure 28. Indoor LiDAR 3D perspective plot examples
Figure 29. Archaeological Site LiDAR scan, sampled data, and labeled reference groups 109
Figure 30. Archaeological Site LiDAR segmentation performance curve examples
Figure 31. Archaeological Site LiDAR segmentation surface plot examples
Figure 32. Archaeological Site LiDAR 3D perspective plot examples

KEY TO SYMBOLS OR ABBREVIATIONS

3D	Three-dimensional
2D	Two-dimensional
AB	Point values for the CIE La*b* color space, without Lightness
AUC	Area under the curve index
Avg. AUC	Average area under the curve index of Mean Shift bandwidths
FAMS	Fast Adaptive Mean Shift clustering algorithm
GPS	Global positioning systems
LiDAR	Light Detection and Ranging
LAB	Point values for the CIE La*b* color space
LUV	Point values for the CIE Luv color space
Max. AUC	Maximum area under the curve index of Mean Shift bandwidths
Mr	Correct segmentation metric for reference groups
Ms	Correct segmentation metric for segmented output groups
Mro	Over-segmentation metric for reference groups
Mso	Over-segmentation metric for segmented output groups
Mru	Under-segmentation metric for reference groups
Msu	Under-segmentation metric for segmented output groups
MV	Point values for mean vector variable in the x, y, and z, axes
MVx	Point values for mean vector variable in the x axis
MVy	Point values for mean vector variable in the y axis
MVz	Point values for mean vector variable in the z axis

- OAM Overlapping Area Matrix
- RGB Point values for the red, green, and blue color space
- UV Point values for the CIE Luv color space, without Lightness
- XYZ Point values for the x, y, and z axes

Chapter 1 - INTRODUCTION

1.1 Major Concepts

The study of Geography generally concerns phenomena at or near the surface of the earth. With the advent of technologies such as remote sensing, humans have boosted their ability to sample and measure characteristics of this environment. We are now able to sample vast regions of previously inaccessible locations on the earth's surface as well as in the atmosphere and below the surface of land and water bodies. Higher resolution data are also being collected as sensors and computer processing power are improved. Due to these new capabilities, geographical data sets have become very large. With more geographical characteristics being sampled at higher resolutions, it is difficult to ascertain the general geographic phenomena through their data representations. The data are qualitative or quantitative values of some characteristic at particular locations in space and time: a finite sampling of locations representing geographic phenomena. They can be grouped to reveal representations of contiguous regions of spatial and thematic homogeneity, either parts of or whole geographic phenomena.

My thesis is concerned with finding groups of data representing geographic phenomena in the environment: specifically, groups of 3D points with similar locations, spatial relationships, and thematic values of spectral reflectance. Major relevant concepts include the representation of geographic features as objects or fields, cartographic generalization, and geographic ontology. I rely on the subfields of GIScience, Cartography, and Remote Sensing from the discipline of Geography; Image Analysis and Computer Vision from the discipline of Computer Science; and Cluster Analysis from information sciences. The main topics I incorporate are representation, sets, generalization, aggregation, classification, segmentation, and clustering.

Following this Introduction, I discuss these concepts and topics in my Conceptual Framework section. Implications and decisions relating to my experimental design are also covered in my Conceptual Framework. The Methodology section involved details of my actual experimental procedures as well as two case studies. The Results and Discussion and the Conclusions follow.

1.2 Focus of Work

As an introductory preview, I now briefly describe my topics. Later, in the Conceptual Framework section, I will discuss these topics in detail. Figure 1, is a general graph summarizing the theory, processes, methods, and results of my work, excepting the experimental accuracy assessment. The next few paragraphs elaborate on these components.



Figure 1. General summary of theory, processes, methods.

Representation of geographic phenomena is my main concern. I employ a three-dimensional (3D) point data structure. From the GIScience literature, I recognize these 3D spatial locations are samples of phenomena that could be considered either objects or fields. The points are considered unstructured in the sense that there are no topological connections or relationships initially defined. As 3D points, they could represent locations on surfaces or throughout volumes. Spatial location is an inherent property of the samples. However, I also developed a measure of each point's spatial relationship to its neighbors to identify similar groups of points representing particular orientations or spatial arrangement of the geographic phenomena.

The point locations also have thematic attributes associated with them. The thematic attributes could be any quantitatively measured or qualitatively labeled characteristics of some geographic object or field phenomena; I selected spectral reflectance. In proposing groups of data points that represented some geographic phenomena, two important considerations arose. First, the concept of spatial autocorrelation was relevant as characteristics of nearby locations are more likely to be similar than distant locations. This supports the idea that spatial location itself is a primary indicator in finding locally similar points. Contiguous sets of similar points are further defined with respect to their spatial relationships or thematic characteristics. Secondly, local groups of points can, to some degree, be employed to predict the thematic character of unsampled locations within the region they encompass.

Keeping focus on what such groups of points represent: geographic phenomena. I am interested in synthesizing elements of two geographic theories. I notice similarities between cartographic

generalization and geographic ontology. A particular process of cartographic generalization known as aggregation is used to group points that represent homogeneous regions. Likewise, in geographic ontology, low-level observable points are grouped into sets representing simple objects. This shows conceptually parallel processes, both developing regions of similarity from groups of points.

Grouping sets of data by measures of similarity is an initial step in the common process of classification, before the sets are actually labeled as belonging to a class. I am interested in grouping 3D point data into locally homogenous groups representing geographic phenomena. Finding component parts of objects or geographic phenomena with many smaller groups is preferable to finding too few groups which do not adequately separate the geographic phenomena. The component parts can be aggregated later, if necessary.

As described in the Methods section, my experimental design used synthetic 3D point data sets with various spectral values as thematic attributes. I employed the Mean Shift clustering method for grouping points based upon their spatial location, relationships, and thematic attributes. Along with a brief comparison to a K-means clustering, I extended an accuracy assessment framework to quantitatively evaluate the clustering performance for synthetic data and two case studies.

1.3 Contributions

This is a summarized list of my research contributions. Details of the topics, implications, and experimental decisions are covered in the Conceptual Framework section. The Results, Discussion, and Conclusions sections also add insights related to these contributions. This listing

is loosely based on the order of introduction of the topics with earlier topics building a foundation for subsequent topics. Here are my contributions to the discipline of Geography:

1) My research utilized a <u>3D point data</u> structure. Though relatively uncommon in Geography, this type of data is becoming more prevalent as it offers more flexibility for complex geographic representation than the historically common 2D point or 2D raster data.

2) I developed a measure of spatial neighborhood relationship for unstructured 3D points called the <u>Mean Vector</u>. The name refers to the process of finding the (average) mean of 3D vectors from each point to its local neighbors.

3) For thematic attributes, I evaluated the effectiveness and potential benefits of three common *color spaces* (RGB, LUV, and LAB) for grouping 3D points.

4) I employed the <u>Mean Shift clustering</u> technique which is rarely used in geographical data analysis. It is a successful segmentation technique commonly used in Computer Science for 2D image analysis and computer vision.

5) I employed a remote sensing *segmentation performance accuracy assessment* framework developed for 3D range scans.

6) I extended the accuracy assessment to include numerical *area under the curve indices*.

7) I developed new <u>segmentation performance surface plots</u> for comprehensively evaluating three dimensions for my segmentation performance assessment (tolerance, bandwidth, and instances of correct segmentation).

8) Theoretically, I combined elements of two geographical concepts, <u>*cartographic</u>* <u>*generalization and geographic ontology*</u>, formalizing a particular process and method of geographic representation.</u>

1.4 General Hypotheses

My first general hypothesis is that groups of data representing geographic phenomena in the environment can be identified using spatial location, spatial relationship, and thematic characteristics with the Mean Shift clustering technique.

My second general hypothesis is that combinations of up to three variable sets (spatial location, spatial relationship, and thematic attributes) enhances the ability to find groups of data representing components of geographic phenomena, as compared to single variable sets.

The next chapter describes the conceptual framework in more detail. Chapter three presents the experimental methods. Chapter four reports the results and includes discussion, and chapter five concludes with some general thoughts.

Chapter 2 - CONCEPTUAL FRAMEWORK

2.1 Overview

The Conceptual Framework section discusses previous research and implications to the current work. My experimental design decisions and innovations are also included. The chapter begins with a discussion of theoretical viewpoints on geographical representation and data models. Then, particular topics relating to cartographic aggregation and geographic ontology are introduced. From these two fields of study, I draw a parallel between two similar conceptual processes and discuss a theoretical synthesis. I then provide conceptual background on the methods. First, classification and multivariate clustering methods are introduced. Next, the Mean Shift clustering is described. Finally, an established accuracy assessment framework is both presented and extended. This Conceptual Framework introduces and explains portions of the methods, reducing lengthy discussion in the following chapters.

2.2 Geographic Representation

The study of geographic features involves at least two components, spatial location and thematic attributes (Goodchild, 1990). Data for analysis are obtained by sampling features in the environment. The data are never as complete or as detailed as the feature being sampled. Therefore, geographic data are always generalized representations of some environmental phenomena with the information often presented in maps. When the data are entered into a computational environment, models are used for representation and prediction of some phenomena in geographic space (Frank, 2000).

A review of geographic representation and many types of data models (Pequet, 1988) described a dual identity for maps: the data are represented as either as images or as geometric structures in

graphic form. Many versions of raster models, for images, and vector models, for objects, have been employed for analysis in computing environments (Pequet, 1988).

Subsequent academic work on the dual representation models has argued that there is a complementary role that the two models offer for representation of features in geographic space (Couclelis, 1992). The work of Couclelis (1992) described the vector data as objects, treating geographic phenomena similar to geometric shapes or table-top items that can be manipulated. Couclelis (1992) addressed the common challenge of region delineation, and noted that the majority of hard boundaries were related to human artifacts such as engineered structures and administrative boundaries. The use of vector structures for geographic objects also provides benefits such as the potential for specification and analysis of spatial and topologically connected relationships. Couclelis (1992) also pointed out that raster data usually represents geographic phenomena as fields. Groupings of clustered pixels that share particular values of some thematic attribute can be identified as features. However, the boundaries of such regions are not as discrete as in the object model. The field model described by Couclelis (1992) also allowed the possibility of different features being found in the same location if different thematic attributes are observed. Geographic features could be represented with either model, though the model should be chosen to best fit the geographic phenomena or properties being studied.

A recent article (Goodchild, Yuan, & Cova, 2007) formally synthesized the two data models by first considering the previous discussions (Pequet, 1988; Goodchild & Wang, 1989; Couclelis, 1992) and distinctive perspectives of geographic phenomena. The object-view perspective characterized geographic space as mostly empty and containing objects discretely defined as

points, lines, polygons, or volumes in two, three, or 4 dimensions. The field-view perspective characterized geographic space as filled with variable values of some measureable continuous phenomena. The synthesis of these two representational models was accomplished through reference to the nature of the geographic feature of interest. The shared foundation of both models was termed the atomic form of geographic representation (Goodchild, Yuan, & Cova, 2007). The first primitive of this representation, the geo-atom, was defined as the association between a point in space-time and some property, or thematic attribute. In essence, there may exist an infinite number of infinitesimal (vector) point locations associated with some measured value of any (field) property. Geographic features represented by many (sampled) points of similar values for the particular property may be considered either an object or a region within a continuous field.

For my thesis, the theory of atomic geography helped to connect the vector model of point data with the regions of similar spatial and thematic properties that they represent. My synthetic data sets, for controlled experiments and accuracy assessment, consisted of points representing simple shapes with very separable color attributes. Points were configured as groups, each having similar values, intended to represent individual objects or components of objects. The experimental methods were then employed with the data sets of my case studies to demonstrate the potential for identifying regions of homogeneity in the environment. These data were sampled with a terrestrial LiDAR sensor and represented real-world scenes having complex spatial and thematic properties. They were discrete samples of a particular point in 3D space and a measureable thematic property (spectral reflectance). The vector point data were entered into a computational environment to identify groups of points representing regions of similar spatial

and thematic properties. The theory of atomic geography provided a link between the vector point data model and the 3D features the points represented in a continuous field of measurable spectral reflectance.

2.3 Cartographic Generalization

In Geography, paper maps have traditionally provided a generalized representation of geographic phenomena. A relatively recent comprehensive theory of generalization was presented by McMaster and Shea (1992). Beginning with a review of definitions and numerous theoretical frameworks which preceeded their work, McMaster and Shea (1992) ultimately offered a theory of cartographic generalization suitable for computational use in a digital environment. Of the main reasons stated for generalization was the aim of reducing complexity. As such, generalization essentially reduced the amount of application-specific data and detail to retain clarity of a map, especially when the map scale was reduced.

Of particular importance for this thesis is the generalization of point objects representing regions of similar spatial and thematic properties. McMaster and Shea (1992) define the term *aggregation* as the spatial transformation operator which aggregates, or groups, (nearby) points having similar attributes and represents them as polygonal area features. They note a dimensionality transformation from a point representation to an area representation. They also reference Keates' (1973) example of aggregating individual buildings which are near to each other and representing them as built-up areas. This thesis is concerned with aggregating individual (synthesized or sampled) 3D points into groups which represent regions of spatial and thematic similarity.

Though this thesis is mainly concerned with representation of geographic features, the methods employ computational processes. Weibel (1997) discusses two views of generalization: representation-oriented and process-oriented. Representation-oriented generalization is concerned with the representation of particular features at different map scales. He mentions that a multi-scale database may be required to provide different sets of information at different scales, and if alterations to the data are made at one scale, they may not be automatically updated at the other scales. The process-oriented view is the focus of Weibel's (1997) work. This is described as being more automatic and flexible as all of the map scales are derived from a single spatial database using generalization algorithms. As detailed in my methods, I employ semi-automatic algorithms that only require certain input parameters: number of output groups for k-means clustering, and number of neighbors and bandwidths for Mean Shift clustering. These belong to the global and bandwidth classes, respectively, of spatial data generalization algorithms described by Weibel (1997).

Weibel (1997) mainly contrasts the expert-driven, creative techniques of knowledge-based cartography with the use of generalization algorithms. However, he notes that they can be used together for improved results and that both approaches still require key decisions by the user. Human input guides cartographic generalization with either subjective and intuituve elements or with control of algorithmic processes. He also implies that well-defined objectives improve the result of either approach.

2.4 Geographic Ontology

To produce representations of geographic features from point data with 3D coordinates and thematic attributes, certain objectives are required. In other words, a conceptual model, or

ontology, for the features of interest must be defined. Ontology, as a philosophical discipline, is concerned with conceptualizations or particular categories for particular views of the world (Guarino, 1998). The concepts are ideas of things in existence depending upon particular perspectives. Another related meaning for the term ontology has become popular due to developments in information systems and computer sciences. Guarino (1998) mentions that this meaning generally refers to domain modeling as implemented by various methodologies. In geographic information sciences, ontology also relates to the specification of a conceptualization and both lines of investigation are pursued: top-level categories from a formal perspective, and a domain-specific ontology from a task-oriented approach (Agarwal, 2005).

I followed a particular domain ontology in this research. In practical terms, this can be generally described as knowledge representation, defining features as components of the concept (Schwering, 2008). My ontology related to both the conceptual categories and physical properties of spatial and thematic data representing 3D features. I aimed to find components of objects in a scene that have similar spatial and thematic characteristics.

Features in a landscape scene can become more distinctive when additional information is added. Couclelis (2009) recently described 7 levels of feature representation depending upon the amount of semantic information attached to a feature and its components. A description of relevant portions of her hierarchy follows. At the lowest ontological level, only points in space-time (locations) relating to a particular purpose exist. Adding crude, qualitative knowledge at the point locations provides the next higher level of information, as (thematic) observables. Separating the points into classes, based upon the observable qualities, provides additional another level of (generalized) information. Then, spatially connected homogeneous regions are identified as simple objects (real-world entities). Assemblages of simple objects can be associated, providing what she terms as complex objects (this is one level beyond my methods, but included in my discussion.) I follow her hierarchical levels of features and their components.

The three types of characteristics which made up my ontological framework were: spatial location, spatial relationship, and thematic attributes. The data employed in this thesis initially had 3D coordinate values for the x, y, and z axes as well as red, green, and blue color values. These basic characteristics were used with additional variables calculated from this data: a measure of spatial relationship and additional color spaces, each described in more detail later. The data values were expected to differentiate separate groups of data points representing different components of the 3D scene. As exemplified in the following three paragraphs, I intended to identify components that:

- were in different locations (in 3D space)
- had differing spatial relationships to neighboring points (planar or non-planar)
- had different thematic attributes (colors)

Spatial location can be a very useful characteristic to distinguish different features having similar thematic attributes or local spatial relationship. For example, walls of separate buildings may have a similar color and (a locally planar) spatial relationship, though we generally consider them as components of different buildings based upon their location. The same could be said of similarly shaped and colored groups of vegetation in different locations.

Local spatial relationships could provide an additional set of information, separating features having similar colors and being in close proximity. Examples of objects that could be separated based on (locally planar) spatial relationships include a concrete sidewalk that is adjacent to gray boulders or a grassy area that leads to green bushes.

Thematic attributes, such as spectral reflectance in the red, green, and blue wavelengths, can be used to separate nearby regions with the same spatial orientation. For example, a study on the percentage of impermeable ground cover in an urban scene may have data representing an asphalt roadway, a sidewalk, and grassy landscaping. All of the features may be relatively flat and level. Color can be used to separate the regions.

My aim is to identify features in a scene represented by 3D points by using the three basic characteristics of location, spatial relationship, and spectral values. The resulting aggregated groups of points may represent simple objects that could be subsequently assembled into complex objects, if necessary. The generalized representation and separation of groups with distinctly different characteristics are expected to ease interpretation of complex 3D data and landscape scenes.

2.5 Synthesis of Cartographic Generalization and Geographic Ontology

This thesis closely followed two theories: the cartographic generalization of McMaster and Shea (1992), and an ontological hierarchy from Couclelis (2009). These two theories, though they came from separate threads of geographic inquiry, are very parallel in relation to this investigation. Both are related to geographic representation of features in the landscape and both present processes that similarly combine information to better understand those features. The

generalization process of aggregation was a transformation that grouped 3D points into representations of areal regions having similar spatial and thematic characteristics (McMaster & Shea, 1992). The ontological hierarchy also developed simple objects from spatially connected, homogeneous space-time points with similar observables (Couclelis, 2009). These two theories eased interpretation of complex data sets by providing a theoretical basis for the development of groups of points that represent recognizable features in the landscape.

2.6 Spatial Characteristics

The initial data included 3D spatial location coordinate values in the x, y, and z axes (XYZ). Though spatial location was the lowest-ontological level of information, it provided a measure of proximity between points and group points representing contiguous regions.

Ledoux and Gold (2008) argue that processing objects in 3D may have advantages over common raster computations, despite the ease of computing with raster arrays. Spatial data consisting of x, y, and z values can be sampled; they maintain natural discretization; map algebra operations are still available; and there are potential benefits for analysis and visualization stages (Ledoux & Gold, 2008). One of the first steps in using 3D point data is to identify features in the scene for further quantitative analysis and qualitative evaluation.

2.6.1 <u>Derived Spatial Variables - Neighborhood Relationships</u>

To augment the available variable sets, I propose a metric for 3D spatial relationships intended for identifying planar regions. It is calculated for each point using the relative locations of neighboring points. With 2D raster grids, the slope, curvature, and roughness could be calculated from adjacent cells and used as neighborhood characteristics. Similarly, if a digital surface model was already created from the 3D data points, the topological structure and connectivity could be used to determine which points were adjacent to calculate similar measures as in the 2D raster case. However, the goal of my thesis is to identify groups of unstructured 3D points where no topological connectivity between the points is defined. By identifying groups first, interesting data can be selected before spending the time and computing power on such processes as surface interpolation. This also allows the method to be more generalizable and applicable to any 3D spatial data, representing surfaces or other phenomena. Therefore, a new measure of spatial neighborhood relationship is considered: the mean vector (MV). This is simply the mean of vectors from each point to its neighbors, as defined in the Methods section.

The neighboring points in the MV calculation could be determined in several ways: number of neighbors, distance with an absolute radius, distance with an adjusted radius based upon local density, or even a hybrid method of sampling points within a radius. For simplicity, I chose the number of neighbors method. I decided 8 neighbors would be beneficial as a small set of 8 neighbors helps to boost differentiation between local regions by not averaging across too many points. It would also be computationally efficient and is somewhat comparable to the queen's case of neighbors in a 2D grid.

I expect that the MV would help to group 3D points representing planar regions as opposed to edges, corners, or regions of high variability. Figure 2 helps to illustrate the following example. First, I'll describe the situation considering points toward the interior of the planar region, away from edges, on the top side. The vectors to its neighbors would not vary in the Z axis, the X values of the vectors to the points to the left and to the right would offset, and the Y values of the

vectors to the points above and below would offset. This results in a mean vector at or very near zero. Most points on each side would have this same near-zero value for the mean vector. This value will help to distinguish planar regions as being different from edges, as described next.

Points near an edge, or a change in the surface orientation, on one side may have neighbors representing a different side or orientation. When the coordinates from different sides are used for the mean vector calculation, the non-planar neighbors will result in a mean vector that varies in up to three dimensions. The mean vector metric could be considered beneficial for edgefinding or, as in my aims, for separating points with near-planar spatial relationships from points representing edges, corners, or highly variable regions.



Figure 2. Mean vector relationships of 3D points representing an object. (For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis.)

2.7 Thematic Characteristics

By adding measureable values of thematic attributes, the data have a higher ontological level and more information is available for cartographic generalization processes. Studies that use both spatial and spectral properties for segmentation are of particular importance to this thesis, since this is how I aim is to identify components of objects within a 3D scene. For instance, laser scans provide detailed geometry but still require interpretation for mapping and segmentation of a complex scene is not possible with a single spatial cue (Barnea & Filin, 2008).

The previous work by Barnea and Filin shows that segmentation methods using either range or color alone have limitations, but together they have better results. The most promising method they implement, which I will also evaluate, is the use of Mean Shift clustering (Comaniciu & Meer, 2002). I will describe the Mean Shift clustering in the Conceptual Framework section on Clustering. Compared to other clustering methods, it is relatively low-parametric and only requires a bandwidth window of evaluation as described. It can also work with multi-variate data. For their study, Barnea and Filin used range information, surface normals, and color along with an iterative approach to further separate under-segmented clusters. However, to address issues of relatively high LiDAR point density near the scanner, they make a data representation transformation from the 3D object-view (empty space with objects) to a more traditional remote sensing computational model, the 2D field view (a raster grid with all cells filled with a value).

I followed their approach in regard to the combination of spatial and thematic information and the use of Mean Shift clustering. However, I maintain 3D object-view representation of the data, having unstructured 3D points in space with additional thematic attributes.

2.7.1 <u>Derived Thematic Variables - Spectral Color Space Alternatives</u>

The basic synthetic data included a set of red, green, and blue (RGB) values which can be considered a surrogate for any thematic variable. There are alternative color spaces for visible wavelengths which I was interested in evaluating. These were expected to assist in reducing the impact of shadows or environmental illumination on the spectral values. For instance, the the CIE Luv (LUV) and CIE L*a*b* (LAB) color spaces isolate lightness into one variable and have two variables for color. They are also designed to have values that more evenly represent the humanly perceptual differences between colors. Research has deemed CIE L*a*b* as an appropriate color space for raster image segmentation along with the Mean Shift clustering (Paris & Durand, 2007).

I chose to evaluate the original RGB along with the LUV and LAB color spaces. For my experiments with synthetic data I made use of the RGB and all three variables in each of the LUV and LAB variable sets. However, the LiDAR case studies had highly variable RGB values, changing material reflectance, and shadows. To address these issues and improve segmentation performance, I made use of supplementary variable sets, UV and AB, removing the (L) lightness variable from the LUV and LAB variable sets.

2.8 Classification and Segmentation

Using basic characteristics for classification of landscape features is common. The variable spatial and thematic characteristics can be quantitatively evaluated and converted into a generalized representation of classes. Instead of attempting to deal with an infinite variety of variables and values for an entire region, the sampled data can be reduced to a more

understandable set of thematic classes (Foody, 2002). This classification provides another set of semantic information, raising the ontological level up in the Couclelis (1992) hierarchy.

Computational environments have greatly aided the processing of such abundant information. From the computer science literature, Haralick and Shapiro (1985) describe processes such as image segmentation which have created many possible avenues for classifications of land cover, clarifying and aiding in our interpretation landscape features.

Figure 3 shows an example of image segmentation based on classified spectral values. For this classification, a Landsat 7 ETM+ satellite image is used, showing a region in the vicinity of the Michigan State University campus. The first image displays the region in true color, using bands 1, 2, and 3 (blue, green, and red). The second image displays the region in false color, using bands 3, 4, and 5 (red, near-infrared, and mid-infrared). The third image displays the classification results. Pixels are grouped into 4 classes of land cover using all of the spectral values available in the satellite image. This generalization techniques helps to represent and evaluate landscape features.



Figure 3. Segmented satellite image (true color, false color, and classified).

In a theoretical sense, classification is often related to both cartographic generalization and geographic domain ontologies, since observable properties are selected and used to represent features in the landscape. In a practical sense, classification often employs clustering, described in more detail in the next section. This can be implemented in either supervised or unsupervised methods. For the supervised case, data are selected by an analyst to provide training signatures for the desired classes. Each of the remaining data are assigned to the most similar class through a clustering process. For unsupervised classification, no classes are initially specified. Clustering is performed on the entire data set and the resulting groups are labeled by the analyst.

2.9 Clustering

In this section, I discuss multi-dimensional clustering, starting with the general theory and following with specific clustering methods. I present the Hierarchical, K-means, ISODATA, and the Mean Shift types of clustering to show a lineage of methods related to my implementation of the Mean Shift clustering. To offer a basis of experimental comparison, the well-known and popular K-means method is used to evaluate the performance of the Mean Shift method.

In general, clustering can be beneficial to understanding multidimensional data sets. A review of data clustering (Jain, Murty, & Flynn, 1999) states that the aim of data clustering is to determine natural groupings and they explain that this is typically achieved by minimizing within-group variance while maximizing between-group variance. They describe two conceptual approaches to this goal. One approach is agglomerative clustering, adding similar data to a group in sequence. The second approach is divisive clustering, sequentially separating the most differing data apart. Finally, they point out that differing results that can be achieved by clustering. Hard assignment considers the destination cluster for a point to be one single group. On the other

hand, fuzzy probability clustering can allow for a point to potentially belong to one of several groups with differing degrees of probability.

My experiments follow the concepts of unsupervised, agglomerative clustering with hard assignment. Numerous types of clustering are available; a few are described below since they relate to the Mean Shift clustering approach which was ultimately selected for finding groups of points with similar spatial and thematic characteristics.

2.9.1 <u>Hierarchical</u>

The basis of clustering philosophy can be easily conceptualized with a description of hierarchical clustering. Hastie, Tibshirani, and Friedman (2009) explain that hierarchical clustering starts with each and every point in a cluster of its own. Then, the (typically Euclidean) distances in the multi-dimensional variable space from every point to every other point is determined and the two closest points are combined into a cluster. Once they are clustered, the values of the points will be averaged to create a new value for the center of the cluster. This continues iteratively until all of the clusters have been combined into one group. A dendogram is created and used to select a cutoff point for a reasonable number of clusters.

2.9.2 <u>K-means</u>

The k-means algorithm minimizes the within cluster variability and maximizes the between cluster variability (MacQueen, 1967). The number of seed clusters must be selected *a priori*; a non-trivial drawback, since this parameter is often unknown before hand. Xu and Wunsch (2009) provide a description for the iterative K-means process which can be followed in
Figure 4, below. The K-means clustering starts with (K) randomly placed seed values. In this example, four initial seed locations are shown with black squares. For each point in the data set, the multi-dimensional distances to the seeds are calculated. All of the points closest to each seed are grouped together into clusters, as shown in the lower images with group boundary lines. For each group, the mean is calculated, creating new cluster centers. Then, for the next iteration, new groups of points are selected as being closest to the cluster centers and new means are calculated. This continues iteratively as the groups minimize the within-cluster variance and maximize the between-cluster variance. The process tends to produce hyperspherical clusters and will have different results with different initial seed values or different initial seed locations.



Figure 4. K-means clustering process.

To demonstrate the effect of different seed values on K-means clustering, a second iterative sequence is shown in Figure 5. From the previous example, the same set of data points and only

three of the four original seed values are used. This represents a mismatch between number of data clusters (four) and number of seed values (three). By selecting a different set of seed values at the onset, the results are considerably different. All of the resulting groups are different than the previous example, but the most notable difference is that the cluster in the top left was effectively split and combined into nearby clusters. This shows that, by selecting the different numbers of desired clusters for the K-means method, the results may not reflect natural groupings in the data.



Figure 5: K-means clustering mismatch between 4 clusters and 3 seeds.

2.9.3 <u>ISODATA</u>

An example of numerous alternatives related to K-means clustering is ISODATA (Iterative Self-Organizing Data Analysis Technique) (Ball & Hall, 1967). Though the two algorithms are similar, a distinct difference is that the K-means assumes the number of clusters is specified *a* *priori*, while the ISODATA algorithm allows for the adjustment of different numbers of clusters based upon a specified threshold. The original cluster locations are devised by calculating the mean of all variables, expanding the range of cluster origins to one standard deviation, and finally dividing the range per number of cluster origins for starting positions (Tou & Gonzalez, 1974). Though the ISODATA method is not compared here, it shows that the K-means has inspired new ways of finding natural groupings of data. This provides support for methods moving beyond the K-means and is an example of clustering techniques that do not require the number of groups in a scene to be known *a priori*, like the Mean Shift method discussed next.

2.9.3.1 Mean Shift

The Mean Shift clustering (Comaniciu & Meer, 2002) is again similar to K-means. However, the number of output groups is not required. Instead the Mean Shift uses a bandwidth distance, or window of evaluation, for calculating local density. First, windows are placed at either uniform or selected locations, depending upon the implementation. Then, for all of the points within the window, the mean for each variable is determined. A vector from the original location to the mean location is calculated and the window is shifted to the mean, or center of local density. This continues iteratively, calculating new means and shifting the window until a specified convergence threshold is reached and the shift is considered small enough to terminate. It is considered a mode-seeking process. Segmentation of multidimensional data sets with the Mean Shift is achieved by placing many windows for the mean shift process, possibly at data point locations, and grouping data that converged to the same local density values, or destination "modes."

Figure 6 illustrates an example of the Mean Shift using the U and V dimensions (from the LUV color space) of one of my case studies (the Indoor LiDAR data, described in the Methods section). The first of the three images shows the two variables spatially mapped with their U and V values (in standard deviation units) and colored with their RGB values for reference to the real world objects they represent. The second and third images are color and grayscale examples of the same Mean Shift calculations. Both images have 5 selected points with a circle showing the first bandwidth window of evaluation. Here, the bandwidth is 0.5 standard deviation units. The iteratively calculated local density means are also shown, demonstrating 5 paths that lead to 5 different cluster centers. For segmentation using the Mean Shift clustering technique, all data points will be evaluated in similar manner.



Figure 6. Mean Shift on 2D point set, finding modes.

To extend the example to multidimensional (3D) space, Figure 7 shows the three LUV variables spatially mapped to the x, y, and z axes. To aid visual reference of the real objects they represent (in the Indoor LiDAR scan), the data are colored with their associated RGB values in the first image. The second image shows the Mean Shift clustering results. After all of the data points have been evaluated with the Mean Shift clustering, they are colored per their output cluster

groups. Points that arrived at the same local density point have the same color. Even though a few points were not found to be associated with clusters when using the 0.5 standard deviation bandwidth, the Mean Shift clustered the majority of points into 6 clusters without the need for specifying the number of clusters.



Figure 7. Points in a 3D LUV color space clustered into 6 groups by the Mean Shift.

The performance of the Mean Shift has been well documented and extended in the image segmentation literature. For instance, Shimshoni, Georgescu, and Meer (2006) have implemented a version to improve speed by using locality-sensitive hashing to partition the data and reduce the number of iterations needed to find approximate cluster centers.

The Mean Shift has also been applied as a means in quantifying vegetation in an agricultural setting by using raster image color values (Zheng, Zhang, & Wang, 2009), in shoreline extraction using geometry derived from aerial LiDAR (Lee, Wu, & Li, 2009), and in object extraction of urban features from aerial LiDAR points (Yao, Hinz, & Stilla, 2009).

My thesis bridges these studies by implementing the Mean Shift clustering using both spatial and spectral values. As previously mentioned, the spatial location, spatial relationship, and thematic characteristics relate to low-level ontological components in a landscape. With the Mean Shift, the 3D points can be aggregated into locally similar groups for a quantitative and qualitative evaluation.

2.10 Accuracy Assessment

2.10.1 Types of Clustering Assessment

My thesis is concerned with finding groups of 3D points with similar spatial and thematic attributes, possibly representing components of objects. Considering this is multi-dimensional, I have chosen to proceed using clustering aiming to identify groups of similar 3D points. This is essentially a form of unsupervised segmentation of 3D points. Range image classification and segmentation performance assessments offer a means of evaluation. I'll first describe the general types of clustering and segmentation assessments.

A survey of methods for evaluating image segmentation processes (Zhang, 1996) shows there are two main categories to consider. Analytic methods are for evaluating the algorithms themselves; the concepts, principles, and characteristics that are used. Empirical methods, on the other hand, evaluate the performance of the processes through the results. This type of evaluation is very common and can be found to have different branches of study as well. One branch of empirical evaluation is concerned with goodness measures, leaning heavily upon human intuition of a visually "good" segmentation along with measures of intra-region uniformity or inter-region contrast. Another branch of empirical evaluation concerns measures of discrepancy. This deals with differences between ground truth data and the output. Mis-segmentation of pixels, groups, positions, or a variety of characteristics may be evaluated. I will generally make use of empirical discrepancy methods in evaluating my experimental results and the level of performance for the variables of interest.

One of the most basic evaluations of clustering output is simply the number of output groups closest to a predetermined number of reference groups (Zhang, 1996). A more informative metric is the number of correctly or incorrectly segmented points, usually shown as a percentage. These could be helpful as guidance toward a selection of segmentation parameters to be used for more detailed analysis.

2.10.2 <u>Confusion Matrices - Typical Classification Approach</u>

Traditional classification performance assessments establish an error matrix for evaluation (Foody, 2002). In this section I will describe a typical version for classification (having the same number of reference and output classes.) Then, in the next section, I describe a modified matrix that handles my clustering results (having different numbers of reference and output classes.)

As described by Foody (2002), in a typical supervised classification the desired classes to be found in an image are defined. Pixels in an image are selected as training sets and classified as a particular class by an analyst. Every other unclassified pixel is then compared to the training sets and assigned to the class with the most similar multispectral values. All of the pixels are typically assigned to one of the defined classes. An important point here is that a square confusion matrix is used to compare the same number of classes in both the reference data and the output. Traditional confusion matrices can be used to calculate accuracy metrics (Foody, 2002). The square confusion matrix has columns and rows for the specified classes and conventionally represent the reference and output labels respectively. In the cells are the number of sampled pixels corresponding to each reference and output class. As Foody (2002) describes, the diagonal sum of this matrix represents the total number of correctly classified pixels in every class and the percentage of correctly classified pixels to the total number of sampled pixels offers the Overall Accuracy of the classification. He also notes that two other metrics are available for each class, the Producer's Accuracy and the User's accuracy. The Producer's accuracy is a percentage measure calculated as the number of correctly classified pixels in a reference column to the total number of pixels in that reference pixels that were labeled as some other class. The User's accuracy is a percentage measure calculated as the number of correctly classified pixels in an output row to the total number of pixels in that output row. This shows how well the classification did to avoid errors of commission, or output classes that contain pixels from other reference classes.

Enhancement of the traditional classification accuracy assessment discussed in the literature (Congalton, 1991) compares significant differences in several matrices using kappa and khat metrics. However, with my synthetic data, the number of Mean Shift clustering output groups is not specified and may not equal the number of input reference groups. Not having a square matrix negates the possibility of using the kappa and khat values for determining if a classification is significantly better than another. Nevertheless, the traditional confusion matrix lays a theoretical foundation for evaluating accuracy, though it requires a modification, as described below.

2.10.3 Overlapping Area Matrices - Modified Segmentation Approach

As the number of object components in a landscape scene is often unknown, it is preferable not to specify the number of output groups in advance. This may result in a (rectangular) asymmetric confusion matrix having different numbers of columns and rows. A symmetric, square matrix, only results if the algorithm happens to produce the same number of output groups as there are reference groups. There were typically 6 reference groups (columns) specified for my synthetic data, though the alternative color configuration had 12 reference groups. The output groups (rows) in my experiments had the potential to range from many groups containing one point each at small bandwidths to one group containing all of the points at larger bandwidths.

A modified type of performance matrix, the Overlapping Error Matrix (OAM), was developed for segmentation evaluation (Ortiz & Oliver, 2006). This segmentation accuracy assessment works with unequal numbers of reference and output groups.

As described in the Methods section, I had the benefit of a known ideal number of groups in my synthetic data set. The Mean Shift clustering output allows for a flexible number of output groups to be found. Thus, the confusion matrix consisted of six columns representing reference groups and a variable number of rows representing the output groups. If there were fewer than six output groups, the clustering resulted in some level of under-segmentation. If there were more than six output groups, the clustering resulted in some level of over-segmentation. A hypothetical under-segmented example output is shown in Table 1; it has six reference groups (columns) and three output groups (rows). The lower and right margins are the column and row totals, respectively.

	Ref. Group 1	Ref. Group 2	Ref. Group 3	Ref. Group 4	Ref. Group 5	Ref. Group 6	Total
Output Group 1	5	15	45	55	45	44	209
Output Group 2	0	40	15	4	10	9	78
Output Group 3	59	9	4	5	9	11	97
Total	64	64	64	64	64	64	384

Table 1. Hypothetical overlapping error matrix showing under-segmentation.

2.10.4 Range Data Segmentation - Extended Metrics

To address 3D spatial point data directly, I followed the literature regarding range image segmentation and implemented the accuracy assessment framework by Hoover, et al. (1996) that extended the measures of an overlapping area matrix. In this assessment framework, several metrics were defined and additional stages were used to determine whether every cell in the matrix was initially labeled as at least one of 5 classifications: correctly segmented, over-segmented, under-segmented, a missing (reference) segment, or a noise (output) segment. The metrics basically related to proportions of intersecting cell values (numbers of points) to their reference (column) and output (row) values. If a matrix cell value was a high proportion in both the reference and output groups, it was generally considered correctly segmented. Ultimately, the cells were used to determine the number of output groups that are associated with each type of classification. Tolerance levels, measures of acceptance or rejection with which the proportions of points were evaluated, helped gauge the degree of segmentation performance. The tolerance levels employed ranged from a majority of points (>50%) to a perfect, correctly-segmented group with all points (100%) in an output group comprising 100% of a reference group. An

introductory description of this assessment method follows, including hypothetical examples; a detailed description is presented in the Methods section.

First, the reference and segment metric values, *Mr* and *Ms*, were generated for each cell by calculating the proportions of points in each cell to the column total for the reference group and to the row total for the output group. This was somewhat similar to the Producer's and User's accuracy in a typical confusion matrix evaluation. Continuing with the under-segmented example OAM, shown below as Table 1, the resulting values from this stage are shown in Error! Reference source not found. and Table 3 with the *Mr* being proportions of points per reference side (column) and the *Ms* being proportions of points per output segment (row) respectively. Two cells, shown in bold, illustrate that each pair exceeds the tolerances mentioned in the next section.

	Ref. Group 1	Ref. Group 2	Ref. Group 3	Ref. Group 4	Ref. Group 5	Ref. Group 6
Output Group 1	0.08	0.23	0.70	0.86	0.70	0.69
Output Group 2	0.00	0.63	0.23	0.06	0.16	0.14
Output Group 3	0.92	0.14	0.06	0.08	0.14	0.17
Total	1.00	1.00	1.00	1.00	1.00	1.00

Table 2. Mr values, proportions of points in each cell relative to the reference totals.

	Ref. Group 1	Ref. Group 2	Ref. Group 3	Ref. Group 4	Ref. Group 5	Ref. Group 6	Total
Output Group 1	0.02	0.07	0.22	0.26	0.22	0.21	1.00
Output Group 2	0.00	0.51	0.19	0.05	0.13	0.12	1.00
Output Group 3	0.61	0.09	0.04	0.05	0.09	0.11	1.00

Table 3. *Ms* values, proportions of points in each cell relative to the output totals.

2.10.4.1 Segmentation Performance per Tolerances - Correct Segmentation

For the second stage, and as a means of determining the level of segmentation performance, the proportion values in every cell in both matrices were evaluated against tolerances increasing from 0.5 to 1.0, in specified steps. By having a cell proportion value relative to the column and row totals that is larger than a tolerance value, this pair of reference and output groups was considered an instance of correct segmentation at this tolerance level. If the tolerance level was 0.5, the number of points in the cell are at least a majority of points in both the reference group and the output group. Of course, 50% of the points for such a cell is a relatively low level of tolerance. So, by increasing the tolerance in specified steps up to 1.0, for 100% of the points, it was possible to reveal which pairs of reference and output groups had a high proportion of shared points.

Using the example metrics in Table 1 and Table 3, above, the segmentation performance was evaluated as proportions of points exceeding increasing tolerance levels. Table 4, shows whether two corresponding cell pairs meet the criteria and are considered instances of correct segmentation since the proportions of points in Table 1 and Table 3 are both greater than or equal to the 0.5 tolerance. As shown in Table 4, increasing the tolerance to 0.6 resulted in only one correctly segmented cell pair, shown in Table 5.

	Ref. Group 1	Ref. Group 2	Ref. Group 3	Ref. Group 4	Ref. Group 5	Ref. Group 6
Output Group 1	0	0	0	0	0	0
Output Group 2	0	1	0	0	0	0
Output Group 3	1	0	0	0	0	0

Table 4. Cells with proportions greater than or equal to a tolerance of 0.5.

	Ref. Group 1	Ref. Group 2	Ref. Group 3	Ref. Group 4	Ref. Group 5	Ref. Group 6
Output Group 1	0	0	0	0	0	0
Output Group 2	0	0	0	0	0	0
Output Group 3	1	0	0	0	0	0

Table 5. Cells with proportions greater than or equal to a tolerance of 0.6.

2.10.4.2 Instances of Over-segmentation and Under-segmentation

The third stage of evaluation addressed instances of over-segmentation. This generally proceeded with the same theoretical concept of meeting the criteria of proportions per reference column and output row. However, over-segmentation occurs when one reference segment is represented by several output groups. Here, instead of simply one row, the sum of the cells from every row that are participating in the column was determined and a proportion of the column total is again calculated for the metric *Mro* (over-segmentation metric for reference groups). However, to be

sure that each of these rows was participating more in this column than any other column, each participating row cell must have met the tolerance value with regard to the proportion of points compared to the row total. If this additional criteria is met, the cells for all the participating rows were summed, and a proportion relative to the row total was calculated for the metric *Mso* (oversegmented metric for output segments). Both *Mro* and *Mso* must have then met the tolerance for the cell to be classified as an instance of over-segmentation.

The fourth stage of evaluation addressed under-segmentation. This is where one output group represented several reference groups. It is evaluated in the same fashion as over-segmentation, although it uses several columns to derive the *Mru* and *Msu* metrics (under-segmented metric for reference groups and output segments, respectively).

2.10.4.3 <u>Final Classification</u>

In this framework, each cell could be classified as a member of up to three classes (Hoover, et al., 1996). The fifth stage of evaluation simply determined which classification the cell was mostly participating in by averaging both measures for each of the three classifications of correct segmentation (*Mr* and *Ms*), over-segmentation (*Mro* and *Mso*), and under-segmentation (*Mru* and *Msu*). The highest average value of the cell's classifications (relating to the proportions of a cell's participation in each classification) is selected as the final classification for that cell. In circumstances with equal averages of the metrics, preference is given to correct segmentation, then to over-segmentation, then to under-segmentation. A cell not classified as correct segmentation, over-segmentation, or under-segmentation is labeled as a missing reference in the reference column and also labeled as a noise segment in the output row.

The five potential final cell classifications ultimately determined how many groups belonged to each classification, based on the intersecting cell values representing proportions of the reference and output groups. Of these five classifications, I was mainly concerned with the number of correctly segmented groups as the single measure of segmentation performance.

2.10.4.4 <u>Segmentation Performance Curves</u>

By charting the number of groups for each tolerance value, with the tolerance on the x-axis and the number of correctly segmented groups for each tolerance on the y-axis, it is possible to graphically present the performance of each variable set. The segmentation performance curves of Hoover, et al. (1996) easily allowed for a visual evaluation of the persistence of correctly segmented groups for each variable set as the tolerance increased. An example of a hypothetical segmentation performance curve is shown below in Figure 8, having the number of correctly segmented groups specified for tolerances from 0.5 to 1.0 in steps of 0.1. Also shown at the top, is an "Ideal" value of 6 correctly segmented groups, equivalent to most of my experiments' prelabeled reference groups (only one of my synthetic experiments had 12 groups.) The segmentation performance curve shows that 6 output groups are classified as correctly segmented at the 0.5 and 0.6 tolerance values, meaning 6 groups had a number of points that was at least 60% of a reference group and 60% of an output group. At the 0.7 tolerance level, only 5 groups met the criteria for being correctly segmented (over 70% of the points of both the reference and output group.) Eventually, the segmentation performance curve reveals 2 correctly segmented groups at the 1.0 tolerance level, indicating that 2 groups contained 100% of the points of both the reference group and the output group.



Figure 8. Example segmentation performance curve.

2.10.4.5 <u>Area Under the Curve Index</u>

To obtain a single numeric performance value as a quantitative metric, I extended the accuracy assessment of Hoover, et al. (1996) by calculating an area under the curve index (AUC) so several results could be compared. This was helpful for summarizing the numerous segmentation performance curves generated for the variable sets in my experiments. The highest AUC ranking indicated the highest performing variable set.

Chapter 3 - METHODS

3.1 Methods Overview

My thesis concerns the separation of 3D point data into groups with similar spatial and thematic characteristics. To that aim, my methods were concerned with clustering segmentation performance using 3D point data, different variable sets, and multiple perturbations of the data. The specific processes employed are described below. Images in this thesis are presented in color.

In the context of segmentation, my thesis generally involved the Mean Shift clustering algorithm. Based upon the computer vision and segmentation literature, I expected the Mean Shift clustering to segment a large group of multidimensional point data satisfactorily. A notable benefit of the Mean Shift method is that a selected bandwidth window of evaluation, which is applied to locations throughout the data space, can yield any number of clustered groups, as opposed to setting the number of groups with prior knowledge.

I used the K-means clustering technique, to compare the Mean Shift to a more familiar clustering method. The K-means requires the specification of the number of output groups as a parameter. The number of groups was selected based upon the known number of reference groups. However, it is important to note that the reference groups were selected intuitively, based upon spatial and thematic characteristics. The K-means algorithm has the potential to reveal different types of groups than the intuitively selected reference groups. A qualitative evaluation of the output groups helped to explore this situation, as discussed in the Results and Discussion section.

The synthetic data set initially consisted of 3D points with thematic information. From this, additional spatial and thematic variables were derived as explained in the Data Sources and Processing sections below. Several stages of evaluation were used to address issues individually, as described below.

Multi-variate data analysis, spatial resolution, sensor accuracy, and more complex attributecoordinate associations are all important components for this research. As an introductory comparison of the clustering methods, the first stage of experimental evaluation considered the parameters and results of the Mean Shift and K-means clustering. The second stage compared the effects of higher and lower spatial resolutions. The third stage of evaluation considered signal error by introducing increasing levels of noise in spatial and attribute variables. The fourth stage used an alternative color configuration. All of the stages up to this point were evaluated with an accuracy assessment framework incorporating various quantitative metrics, graphed results, and 3D perspective plots. After evaluating the experimental results of the synthetic data, I applied the concepts and techniques to two LiDAR case studies. These steps are described in the Processing and Accuracy Assessment sections below as well as in the Results and Discussion section.

3.2 Data Sources

Most of the data in this thesis were processed with the open-source software "R" (v. 2.10.0) (R Development Core Team, 2009). I also used a C++ implementation of the Mean Shift clustering algorithm, "Code for adaptive mean shift in high dimensions" (Georgescu, Shimshoni, & Meer, 2003), obtained from the Rutgers University Robust Image Understanding Library. I collected the case study LiDAR data in August of 2009 with an Optech 36D terrestrial LiDAR scanner and parsed those data with the Optech Parser (v.4.3.8.6).

3.2.1 Synthetic Data Sets

A series of synthetic datasets were generated for the first sets of experiments. Each consisted of 3D vector points regularly spaced along the six faces of a cube, 10 units on a side. Thematic information— color—was assigned to each point. The number and spacing of points, spatial configuration of color on the cube, and degree of additive coordinate and spectral noise varied by dataset and are detailed in the following paragraphs. The use of synthetic data provided known reference values for each observation, which is valuable in any validation effort. The design mimicked a basic real-world application of directing a LiDAR scanner at an object with clearly defined facets and attributes. Finally, the relatively small size of each dataset minimized the computational burden, enabling additional experiments and analysis.

The first step in my processing was to create synthetic data sets. The initial synthetic data set represented a cube that had points on each of its six sides in 3D (XYZ) coordinates ranging from 0.0-9.0. Since there is often uncertainty regarding exact boundaries or edges of regions in most measurements of real world phenomena, the synthetic data excluded point representation of edges and corners (values 0.0 and 9.0). This resulted in an 8 x 8 array of 64 points representing each of the six sides, totaling 384 points for the data set. For the accuracy assessment, all of the points were labeled as belonging to one of 6 reference groups representing the sides of the cube (front, back, top, bottom, left, and right).

Thematic properties were represented by different levels of red, green, and blue (RGB) colors. The value of visible color reflectance is often collected by remote sensors, though these data were simply a surrogate for any other spectral wavelengths or thematic information that may be included with 3D spatial data. For this experiment, values for color were added to the points. Groups of points representing each side of the cube had a different level of R, G, and B color values in a range from 0.0-1.0. Three sides had full-intensity red, green, or blue values of 1.0, while three sides had half-intensity red, green, or blue values of, 0.5 as shown in Table 6 below. A 3D perspective of the synthetic cube is also shown in Figure 9, below; it is an oblique view from above the top, front, left corner. These initial data are called the "noise-free" data set from here on.

	Red(R)	Green (G)	Blue (B)
Reference Side 1 - Left	1 (Full)	0	0
Reference Side 2 - Top	0	1 (Full)	0
Reference Side 3 - Front	0	0	1 (Full)
Reference Side 4 - Right	0.5 (Half)	0	0
Reference Side 5 - Bottom	0	0.5 (Half)	0
Reference Side 6 - Back	0	0	0.5 (Half)

Table 6. Color assignment for thematic variables of the initial synthetic data set.



Figure 9. Synthetic test cube from the top, front, and left corner.

The accuracy assessment required reference groups against which the segmentation output could be compared. Since I expected that the spatial and thematic characteristics of each side of the cube would provide sufficiently separable groups of points, each set of observations comprising one face of the cube was also considered a reference group. These synthetic groups are experimental surrogates for any groups of points with similar spatial and thematic characteristics in a real-world data set, potentially representing components of an object.

The data set now included a reference label (Ref.), spatial attributes (XYZ), and thematic attributes (RGB). I expected that the XYZ coordinates would help to group points by their absolute position. This is a very important determinant of a group of spatial data points, particularly for the aim of representing a locally contiguous region of similar spatial characteristics. The thematic attributes of RGB spectral values are a very common set of characteristics that could also help to differentiate local regions of similarity.

3.2.2 LiDAR Data - Case Studies

Beyond my synthetic data experiments, I employed the methods on two terrestrial LiDAR scans collected from observations of real-world objects in order to evaluate its potential for geographic feature segmentation. The first was a scan of simple, constructed shapes and colors represented in an indoor scene (subsequently called "Indoor"). The second was a scan of an archaeological site (subsequently called "Archaeological Site"), which provided rugged shapes and numerous colors of built structures and background landscapes. All three data sets are summarized in Table 7.

Data Set	Purpose
Experimental Data	To provide source and reference data for methodology and
(initial and modified versions)	assessment framework included in this thesis.
	To apply thesis methodology to simple man-made shapes and
Indoor LiDAR Data	colors for discussion purposes.
	To apply thesis methodology to complex structures and
Archeological Site LiDAR Data	background landscapes for discussion purposes.

Table 7. Data sets.

3.3 Processing

3.3.1 Spatial Variables - Neighborhood Relationships

The next step in the generation of my synthetic data set was to derive additional spatial information from the existing basic data. The aim here was to represent the spatial relationship between each 3D point and its neighbors. Therefore, a basic measure of spatial neighborhood relationship was conceived and calculated: the mean vector (MV). This is simply the mean of vectors from each point to some of its neighbors.

To calculate the mean vector, the eight nearest neighbors to each point were determined. The 3D (XYZ) vector was calculated for each neighbor. Finally, the average values of change in each axis were returned as a single set of 3D distances in the X, Y, and Z directions (MVx, MVy, and MVz). The formula for each MV is as follows:

$$MV_i = \frac{\sum i_n}{n}$$

where *i* is the coordinate value for each axis (x, y, and z), and *n* is the number of neighbors (8 in this case).

3.3.2 <u>Thematic Variables - Spectral Color Space Alternatives</u>

The initial synthetic data included a set of basic RGB values. This is one of numerous color spaces that can be used for analysis. As described in the Conceptual Framework, alternative color representations were considered to address issues of varying lightness in real-world data. From the RGB variable set I chose to derive values sets for both Luv (LUV) and CIE L*a*b* (LAB) color spaces, using the R software "colorspace" library(R Development Core Team, 2009).

3.3.3 <u>Standardized variables</u>

The synthetic data set included all of the variables that I intended to use: Ref., X, Y, Z, MVx, MVy, MVz, R, G, B, L, U, V, L*, a*, b*. These variables are shown as "variable sets" in Table 8. Variable sets, variables and descriptions..

Variable Set	V	ariabl	es	Description	
				Reference Side Labels 1-6:	
Ref	Refer	ence (Group	Left, Front, Bottom, Right, Back, Top	
				X, Y, and Z coordinates	
XYZ	Χ	Y	Ζ	(For absolute position)	
				Mean Vector X, Y, and Z coordinates	
MV	MVx	MVy	MVz	(For neighborhood relationship)	
RGB	R	G	В	Red, Green, and Blue color space values	
LUV	L	U	V	CIE LUV color space values	
UV		U	V	CIE LUV color space values (No Lightness)	
LAB	L*	a*	b*	CIE L*a*b* color space values	
AB		a*	b*	CIE L*a*b* color space values (No Lightness)	

 Table 8. Variable sets, variables and descriptions.

The variable sets had various distributions of values and also had different ranges (e.g., 1-9 for XYZ and 0-1 for RGB). To remove the possibility of weighting one variable more than another

due to the magnitude of its values, I standardized each variable by transforming data values into their z-scores using the following formula:

$$z = \frac{x - \mu}{\sigma}$$

where z is the standardized z-score, x is the point value, μ is the variable's mean value, and σ is the variable's standard deviation. Though the synthetic variables did not have normal distributions of values (real-world values may have a more normal distribution), the results of this process expressed the variability in a standard, unit-less way for the subsequent clustering operations.

3.3.4 Variable Sets and List of Experiments

With all of the variables prepared, I was able to enter them as combinations into the clustering processes. The 23 possible combinations of variable sets are shown in Table 9. I used 15 variable set combinations for the synthetic data evaluation and 8 additional variable set combinations for the case studies.

Used for Experimental	Used for Case Studies	Variable 1	Variable 2	Variable 2
Experimental	Case Studies	Variable 1	Variable 2	Variable 2
	Studies		Variable 2	variable 3
Data Sets	oraares			
Yes	Yes	XYZ		
Yes	Yes	MV		
Yes	Yes	RGB		
Yes	Yes	LUV		
Yes	Yes	LAB		
(No)	Yes	UV		
(No)	Yes	AB		
Yes	Yes	XYZ	MV	
Yes	Yes	XYZ	RGB	
Yes	Yes	XYZ	LUV	
Yes	Yes	XYZ	LAB	
(No)	Yes	XYZ	UV	
(No)	Yes	XYZ	AB	
Yes	Yes	MV	RGB	
Yes	Yes	MV	LUV	
Yes	Yes	MV	LAB	
(No)	Yes	MV	UV	
(No)	Yes	MV	AB	
Yes	Yes	XYZ	MV	RGB
Yes	Yes	XYZ	MV	LUV
Yes	Yes	XYZ	MV	LAB
(No)	Yes	XYZ	MV	UV
(No)	Yes	XYZ	MV	AB

 Table 9. Combinations of sets of variables, steps, and sequences.

I conducted 5 general sets of experiments, as listed in Table 10, below. They are described in the

following sections and were compared throughout the assessment.

Experiment Set	Торіс
1	Mean Shift/K-means clustering comparison
2	Spatial resolution comparison
3	Noise evaluation
4	Alternative color configuration
5	LiDAR case studies

Table 10. List of experiments.

3.3.4.1 Experiment Set 1 - Mean Shift/K-means Clustering Comparison

The first stage of evaluation for the synthetic data compared the Mean Shift and the K-means clustering outputs using the noise-free data set for both. Because the K-means algorithm requires the number of output groups as an input parameter, this was specified to match the number of (6) reference groups. This was the ideal set of labeled reference groups that could be found by either clustering method. The k-means clustering algorithm, as implemented through the R software "stats" library(R Development Core Team, 2009), is based upon an algorithm presented by Hartigan and Wong (1979). The Euclidean distance measure was used for the k-means clustering.

3.3.4.2 Experiment Set 2 - Spatial Resolution Comparison

The data sets for the spatial resolution comparison were prepared similarly to the initial noise-free data set, though they had different point spacing. In comparison to the noise-free data set's spacing of 1 unit, the lower resolution set had a point spacing of three units, while the higher resolution set had a point spacing of 0.2 units. These values were selected as the first even-valued multiples below and above the initial spacing of one. These point spacings avoided computational rounding errors. More importantly, each data set could be divided perfectly in half, able to meet an accuracy assessment tolerance value of 0.5, or 50% of points on each side.

The total number of points in the set in the low resolution set, was 24 with (2 x 2) 4 points per side. The total number of points in the set in the high resolution set, was 11,616 with (44 x 44) 1,936 points per side. These are large differences in numbers of points, but it emphasized the effects of under-sampling and over-sampling during the data collection phase of an investigation.

3.3.4.3 Experiment Set 3 - Noise Evaluation

This set of experiments consisted of data sets that had an incremental amount of noise added. Random samples of a Gaussian distribution with a mean of zero and an incrementally increasing value of variance from 0.02 to 0.10, in steps of 0.02 units were added to the initial raw (XYZ and RGB) data values.

3.3.4.4 Experiment Set 4 - Alternative Color Configuration

The data for Experiment Set 4 was intended to help evaluate the clustering ability with a more complex, alternative color configuration. As described in the Conceptual Framework section regarding Ontology, objects in the landscape may be close to each other and have the same orientation, though they may be considered separate components due to varying color. To address such circumstances, this data set was prepared with separate groups of points that only varied in color (RGB, LUV, or LAB) while being in close (XYZ) proximity and having the similar (MV) neighborhood. Each side had a different color configuration having 2 colors per side, as shown in Table 11. Some colors had the same varying levels of RGB as the original data set, 1.0 or 0.5 for each R, G, or B. Other colors were combinations of R, G, and B, with values of either 1.0 or 0.5. To evaluate whether the clustering would create groups on different sides (across an edge of the cube), the configuration was prepared to have situations where two of the halves on different but adjacent sides had the same color (e.g., Reference Sides 1 and 5, Table 11).

	Red(R)	Green (G)	Blue (B)
Reference Side 1 - Left Side, Top Half	1 (Full)	0	0
Reference Side 2 - Left Side, Bottom Half	0.5 (Half)	0	0
Reference Side 3 - Right Side, Top Half	1 (Full)	0	0
Reference Side 4 - Right Side, Bottom Half	0	0	1 (Full)
Reference Side 5 - Front Side, Left Half	1 (Full)	0	0
Reference Side 6 - Front Side, Right Half	0	0	0.5 (Half)
Reference Side 7 - Back Side, Left Half	1 (Full)	1 (Full)	0
Reference Side 8 - Back Side, Right Half	0	0.5 (Half)	0.5 (Half)
Reference Side 9 - Bottom Side, Left Half	0	0	1 (Full)
Reference Side 10 - Bottom Side, Right Half	0	0	0.5 (Half)
Reference Side 11 - Top Side, Left Half	0.5 (Half)	0	0
Reference Side 12 - Top Side, Right Half	0	0	0.5 (Half)

Table 11. Alternative color configuration values.

3.3.4.5 Experiment Set 5 - LiDAR Case Studies

After completing the work with the synthetic data sets, the methods were applied to real-world terrestrial LiDAR data sets. The first real-world data set was the Indoor scan of simple, manmade shapes having very basic and separable colors, though there was certainly more variability in both the spatial and thematic (RGB color) variables as compared to the synthetic data. The scan originally included over three million points and was sub-sampled to a much more computationally efficient set of ~5,000 points by selecting every 512th point in the data set. This resolution still provided plenty of spatial and spectral detail regarding the objects in the approximately 3 x 2 x 1 meter scene. The original 3 million point data set, before sub-sampling, is shown in Figure 10, below.



Figure 10. Indoor LiDAR data set.

The second real-world data set was the Archaeological Site scan of complex built-up and natural landscape features having considerable variability in both the spatial and thematic (RGB color) variables. The scan included a stone-lined and capped water channel (diagonally from the lower-left to the upper-right of the scene), bare soil (in the lower-right), and grassy vegetation (on the upper-left). This allows for the evaluation of spatial position and complex orientation, as well as the spectral values for different materials. The scan originally included approximately 0.5 million points and was sub-sampled to a much more computationally efficient set of ~8,000 points by selecting every 64th point in the data set. This resolution still provided plenty of spatial and spectral detail regarding the objects in the approximately 4 x 3 x 1 meter scene. The original half of a million point data set, before sub-sampling, is shown in Figure 11, below.



Figure 11. Archaeological Site LiDAR data set.

3.3.5 *Clustering*

For the Mean Shift clustering, I used the Fast Adaptive Mean Shift (FAMS) algorithm, as implemented in the C++ programming language by researchers at Rutgers University (Georgescu, Shimshoni, & Meer, 2003). It is based upon previously established concepts from work by Comaniciu and Meer (2002). Though the FAMS algorithm is intended to address computational performance by sampling the data set, I simply ran the algorithm using all of the data points. Thus, only two parameters were required for this implementation of the algorithm: a bandwidth distance and the number of nearest neighbors to use. I initially selected both 8 and 24 neighbors for the FAMS algorithm. Since there was no difference in the number of output groups for initial stages of several variable sets between these two nearest neighbor selections, I maintained only 8 neighbors for the full evaluation of all of the experiments. The more important Mean Shift parameter was the distance selected for the bandwidth of evaluation. I used Euclidean distances initially ranging from 0.0 to 3.9 standard deviation units, in increments of 0.1. After initial results were evaluated, this was reduced to a range of 0.0 to 2.9 standard deviation units because the larger bandwidths yielded only one under-segmented group of all data points, with no benefit arising from the evaluation of bandwidths higher than 2.9 standard deviation units. Therefore, each of 30 bandwidths was used for clustering each variable set. The results were evaluated, in terms of the number of output groups that most closely matched the number of reference data groups.

3.4 Accuracy Assessment

The accuracy assessment of the clustering followed the previously employed concepts and techniques described in the Conceptual Framework (e.g., Hoover, et al., 1996). However, as discussed in detail below, there were some alterations and innovations to address the implementation of the Mean Shift clustering.

3.4.1 Mean Shift Clustering

The implemented FAMS algorithm for the Mean Shift method used the previously mentioned parameters (bandwidths and number of neighbors) to cluster the variable sets providing two outputs. One Mean Shift output is a listing of the destination mode for every point, in terms of each variable. Occasionally, the bandwidth of evaluation may not calculate an exact destination mode; it repeatedly bounces between two very close destinations. A small convergence value is used to stop the algorithm from running continuously in such cases.

The next step in the algorithm is a routine that "prunes" the modes. Because many points have destinations that are very close to each other but not with exactly the same values, this pruning routine determines points that are very close to each other, within a small tolerance distance, and re-assigns them to a "pruned mode". The routine reduces a larger number of modes (with few

points each) to a smaller number of pruned modes with relatively more points. The outcome is the second set of results, the Mean Shift pruned modes.

3.4.2 Pruned Mode Labeling Modifications

The second Mean Shift output, showing pruned mode locations, does not specify which particular points are associated with the pruned mode location; it simply returns a count of the points at that location. Furthermore, the total of points at pruned modes did not equal the total number of points in the data set. Some points were not in close enough proximity to be considered a member of any pruned mode. To determine which points were nearest to each pruned mode location, I post-processed the results. I identified a bandwidth tolerance in the FAMS algorithm used for pruning the modes. This was 3000 in $1/(2^{16})$ units of the data set, which was approximately a value of ~0.045 in each variable's standardized units. To post-process the clustering results, I first found the closest pruned mode for every point in the data set and then evaluated if it was within the bandwidth used for pruning the modes. If it was, then *the point* was labeled as belonging to that particular pruned mode, or output group. However, in comparing the resulting number of groups and points in each output group to the number of pruned modes for each and every experiment, variable set, and bandwidth, I actually found that the numbers of my labeled groups and points within those groups was close to, but not exactly matching the number of groups and points in the pruned modes. Some were a few groups or a few points less. To get closer to the Mean Shift pruned mode output, I expanded the bandwidth for labeling the points to 1.1 units, which matched the numbers of post-processed groups to the number of Mean Shift output groups with only a very small proportion of exceptions between the numbers of points in the corresponding groups. These errors were considered negligible since they were mostly found in very small clustering bandwidths (0.0-0.2 standardized units) which

had very few groups, indicating under-segmentation or missing segments. The only notable exceptions were with some bandwidths around 1.3, usually with the LUV variable set and again involved under-segmentation.

3.4.3 <u>Segmentation Performance Classification</u>

With the points finally labeled as belonging to specific pruned mode output groups, I proceeded with an accuracy assessment that was mainly guided by Hoover, et al. (1996). First, I created overlapping area matrices (OAMs) for every experiment, variable set, and bandwidth. Then, the number of points in each cell was used to derive the proportional metrics per column and row used for correct segmentation, *Mr* and *Ms*, as well as the corresponding metrics for over-segmentation, *Mro* and *Mso*, and the metrics for under-segmentation, *Mru* and *Msu*. The labeling convention for the symbols metrics follow this convention:

M - indicates a *metric*

r - indicates the metric is in relation to a *reference* group (OAM column)

- s indicates the metric is in relation to an output group, or *segment* (OAM row)
- o indicates the metric is in relation to a measure of over-segmentation
- *u* indicates the metric is in relation to a measure of *under*-segmentation

The metrics are formalized after a mention of a modification to the reference column metrics (*Mr*, *Mro*, and *Mru*) and the tolerance values used. The final 5 classifications of corresponding reference and output groups were derived from all 6 metrics, as described in sections that follow.

3.4.3.1 Modification to the Reference Column Metrics

One important modification was made to the calculation of the proportional metrics. The intent of the accuracy assessment described by Hoover, et al. (1996) is to evaluate a data set that had

every point labeled; this was not the case for the Mean Shift pruned modes. Some of the points were left out of the Mean Shift pruned modes and the subsequent point labeling. A column of the overlapping area matrix represents all of the points in a reference group. If a subset of points (the set of points belonging to the pruned mode) existed in a column, it did not actually represent a proportion of the reference group's initial number of points. To correct this issue, the column total was set to the original number of points in the input reference group, usually 64 points (or 1936 points for the high resolution data set and 4 points for the low resolution data set). Then, for the *Mr*, *Mro*, and *Mru* metrics, the proportion of points in each cell was calculated as a proportion of the original number of points in the reference groups.

3.4.3.2 <u>Tolerance Values</u>

As mentioned in the Conceptual Framework section on Accuracy Assessment, the process included the evaluation of the metrics against the specified tolerance values from 0.5 to 1.0 (a 50 to 100% proportion), in steps of 0.1 (10%). If each cell's metrics met the required criteria, it was temporarily labeled in separate matrices for each of three segmentation classifications (correctly segmented, over-segmented, and under-segmented).

3.4.3.3 Correctly Segmented Metrics

An instance of correct segmentation is determined when the reference group and output segment meet two criteria, shown below, with the majority of points defined by the incremental tolerance values:

1) The majority of points in the reference group are labeled as members of a single output segment, calculated with the formula:

$$Mr = \frac{N_{rs}}{N_r}$$

where Mr is the ratio of intersecting points (N_{rs}) to the total number of points in the reference group (N_r).

2) The majority of points in the output segment correspond to a single reference group, calculated with the formula:

$$Ms = \frac{N_{rs}}{N_s}$$

where Ms is the ratio of intersecting points (N_{rs}) to the total number of points in the output segment (N_s).

3.4.3.4 Over-segmented Metrics

Over-segmentation occurs when one reference group is represented by more than one output segment. In this case the number of intersecting points is the sum of intersecting points in all of the participating output segments. The criteria and metrics are similar to the correctly segmented metrics. However, the second criterion has two stages. The over-segmented criteria and metrics are:

1) The majority of points in the reference group are comprised of the majority of points in many output segments ($S_{s1}+S_{s2}+...$), calculated with the formula:

$$Mro = \frac{N_{rS1} + N_{rS2} + \cdots}{N_r}$$

where N_{rS1} , N_{rS2} , ... are the number of intersecting points in S_{s1} , S_{s2} , ... respectively. 2a) There is a majority of points in *each and every* one of the participating output segments ($S_{s1}+S_{s2}+...$) participating in the corresponding reference group, calculated with the formula:

$$Mso1 = \frac{N_{rs1}}{N_{s1}}, \qquad Mso2 = \frac{N_{rs2}}{N_{s2}}, \qquad ...$$

where N_{rs1} , N_{rs2} , ... are the number of intersecting points in S_{s1} , S_{s2} , ... respectively and N_{s1} , N_{s2} , ... are the number of total points in S_{s1} , S_{s2} , ... respectively.

2b) The majority of points in *all* of the participating output segments $(S_{s1}+S_{s2}+...)$

correspond to a single reference group, calculated with the formula:

$$Mso = \frac{N_{rs1} + N_{rs2} + \cdots}{N_s}$$

3.4.3.5 Under-segmented Metrics

Under-segmentation occurs when more than one reference group is represented by one output segment. In this case the number of intersecting points is the sum of intersecting points in all of the participating reference groups. The criteria and metrics are similar to the over-segmented metrics. The under-segmented criteria and metrics are:

1) The majority of points in the output segment are comprised of the majority of points in many reference groups ($S_{r1}+S_{r2}+...$), calculated with the formula:

$$Msu = \frac{N_{rS1} + N_{rS2} + \cdots}{N_S}$$

where N_{rs1} , N_{rs2} , ... are the number of intersecting points in S_{r1} , S_{r2} , ... respectively. 2a) There is a majority of points in *each and every* one of the participating reference groups ($S_{r1}+S_{r2}+...$) participating in the corresponding output segment, calculated with the formula:

$$Mru1 = \frac{N_{rs1}}{N_{r1}}, \qquad Mru2 = \frac{N_{rs2}}{N_{r2}}, \qquad \dots$$
where N_{rs1} , N_{rs2} , ... are the number of intersecting points in S_{r1} , S_{r2} , ... respectively and N_{r1} , N_{r2} , ... are the number of total points in S_{r1} , S_{r2} , ... respectively. 2b) The majority of points in *all* of the participating reference groups ($S_{r1}+S_{r2}+...$) correspond to a single output segment, calculated with the formula:

$$Mru = \frac{N_{rs1} + N_{rs2} + \cdots}{N_r}$$

3.4.3.6 Final Classification

For the final classification, each set of proportional metrics was first averaged for the correctly segmented, over-segmented, and under-segmented classifications as shown below:

Correctly Segmented Average =
$$\frac{Mr * Ms}{2}$$

Over - Segmented Average = $\frac{Mro * Mso}{2}$
Under - Segmented Average = $\frac{Mru * Msu}{2}$

Then, each cell was classified according to the highest value of the averaged metrics. If the averaged values were equal, classification preference was given in the order of correctly segmented first, then over-segmented, then under-segmented. *The corresponding sets of reference groups and output segments were considered an instance of either correct segmentation, over-segmentation, or under-segmentation.*

If any cell was not labeled as either correctly segmented, over-segmented, or under-segmented, it received two classifications: as a missing segment with respect to the reference group, and as a noise segment with respect to the output group.

3.4.4 <u>Segmentation Performance Curves</u>

As described in the Conceptual Framework, segmentation performance curves were generated. They show the number of instances of each classification, evaluated at the each tolerance level. The number of graphs would have been very large for each of the 5 classifications from 9 experiments, with 15 variable sets, and 30 bandwidths—20,250 graphs. Therefore, I chose to overlay the results of 30 bandwidths in each graph, reducing the number of graphs to 675. Each graph for each classification included the segmentation performance curves for every bandwidth plotted in gray and the highest performing bandwidth plotted in black. Also, *I was mainly concerned with instances (paired reference and output groups) classified as correctly segmented.* This selection further reduced the number of graphs to 135; 15 correctly segmented graphs for the variable sets of 9 experiments. Examples of these graphs are shown and discussed in the Results and Discussion section.

3.4.5 Area Under the Curve Index

While a segmentation performance curve is a useful visualization of experimental results, the information content can be condensed to a metric for quantitative comparison. I extended the previous accuracy assessment of Hoover, et al. (1996) by calculating an additional metric. An area under the curve index (AUC) was calculated from the segmentation performance curves, with a range from a minimum of zero to an ideal index value of one. Since the ideal result of correct segmentation was specified as the number of reference groups, the actual results could be measured against this value. For each bandwidth of each variable set for each of the Mean Shift experiments, the AUC was calculated with the following formula:

$$AUC = A * \frac{Th}{Tl} * \frac{1}{n}$$

where AUC is the area under the curve index for each bandwidth of each variable set of each experiment, A is the area under the segmentation performance curve (calculated with the trapezoidal rule), Th is the high tolerance value, Tl is the low tolerance value, and n is the number of reference groups. The typical values for the formula are shown for 6 reference groups below:

$$AUC = A * \frac{1.0}{0.5} * \frac{1}{6}$$

The AUCs of different experiments provide information for a quantitative comparison of the performance of different variables for segmenting observations. I employed two comparative approaches with the new AUC information. The first approach uses the maximum AUC value of all bandwidths for each variable set and experiment (called "Max. AUC" from here forward). The Max. AUC values were ranked to reveal the highest performing bandwidth(s) and variable set(s) for each experiment. The second approach uses the average AUC of all bandwidths for each variable set and experiment (called "Avg. AUC" from here forward). The Avg. AUC values were ranked to reveal the highest performing bandwidth(s) for each variable set and experiment (called "Avg. AUC" from here forward). The Avg. AUC values were ranked to reveal the highest performing variable sets (over all bandwidths) for each experiment. The AUC rankings are shown in the Results and Discussion section.

3.4.6 <u>Segmentation Surface Plots</u>

Though the segmentation performance curves and the AUC Rankings are helpful aids to understanding the performance of the variable sets, they both failed to provide all available dimensions of performance. The segmentation performance curves lacked a clear view of each particular bandwidth and the AUC Rankings lacked a visual indication as to the character of a variable set's performance over all of the bandwidths. A better way of visualizing the results was necessary. In addition to the other assessments, I produced an innovative representation of the Mean Shift clustering results. It is an extension of the segmentation performance curves for each variable set. I used a common raster surface plot, showing the tolerance values (on the x-axis), the Mean Shift clustering bandwidth (on the y-axis), and the number of correctly segmented groups (as a classified color) for each bandwidth and tolerance, totaling three dimensions. The number of correctly segmented groups is simply represented with color patches ranging from white, for the minimum of zero, to black, for the maximum ideal number of 6 groups (or 12 groups for the alternative color configuration). The segmentation surface plot visualization revealed which variable sets and which bandwidths perform well for each experiment. The Max AUC was readily apparent in many variable sets, bandwidths, and tolerance levels. However, this visualization particularly helped to identify which variable sets performed well over many bandwidths and tolerance levels. It visually demonstrated the characteristics and nuances of the Avg. AUC metric. The segmentation surface plots are shown in the Results and Discussion sections.

3.4.7 <u>Statistical Analyses</u>

Even though the segmentation surface plots and the Avg. AUC showed potential for indicating the persistence of the variable sets' performance over all bandwidths, a more rigorous means of determining the variable sets' performance was through statistical analyses. Several statistical tests were conducted to more precisely understand the relative performance of variable sets. The noise evaluations offered an opportunity to generate enough samples for statistical testing. These multiple realizations of data also allowed for the evaluation of particular variable sets, addressing objectives relating to the most influential color spaces and overall characteristics: spatial

location, spatial relationship, or thematic attribute. The following paragraphs describe the information used and the statistical tests employed: one-sample t-tests, analysis of variance tests (ANOVA), and Post-ANOVA tests. The output is shown in the Results and Discussion section.

3.4.7.1 One-sample t-tests for the Noise Evaluation

To statistically evaluate the resulting effects of increasing levels of noise in the raw data, 19 realizations of each level of noise were generated. The 5 levels of noise were simulated by adding to the noise-free data set individual samples from random distributions with means of zero and a variances ranging from 0.02 to 0.10 units, in increments of 0.02 units. The 19 realizations for each 5 levels of noise were sufficient to run 5 separate one-tailed, one-sample t-tests, comparing each level of noise to a single mean value from the noise-free data set. The Mean Shift clustering results from each variable set was tested with 5 levels of noise. The null hypothesis is that the with-noise data sets have a mean that is equal to the noise-free data. The alternative hypothesis is that the with-noise data sets have a mean that is less than the noise-free data. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The noise evaluation hypotheses were:

Ho: the mean of the with-noise results = the noise-free result

Ha: the mean of the with-noise results < the noise-free result

3.4.7.2 The ANOVA and Post-ANOVA Tests for All Variable Sets

I employed an ANOVA test to compare multiple variable sets and determine if the means are all equal or if at least one of the means is significantly different. On the assumption that some noise is commonly present in most observations and data collection, I used the 19 iterations of data containing the 0.02 level of noise. The Avg. AUC of all variable sets were submitted to the ANOVA test. The hypotheses were: Ho: the mean of variable set 1 = the mean of variable set $2 = \dots$

Ha: not all of the means are equal

Since this only offers an indication of whether there is a difference of means, it was necessary to follow-up with a Post-ANOVA test to determine which means were significantly different from each other. Three Post-ANOVA tests were conducted: the Dunnett-Tukey-Kramer (DTK) test, the Tukey Honestly Significant Differences (HSD) test, and the Least Significant Different (LSD) tests. The outcomes are shown in the Results and Discussion section.

3.4.7.3 The ANOVA and Post-ANOVA Tests for Color Variable Sets

To determine which color space had the highest performance and the most correctly segmented groups over all the bandwidths, an ANOVA test was run solely on the RGB, LUV, and LAB variable sets. The tests used similar data, information, and procedures as mentioned above for all variable sets: the Avg. AUC of 19 iterations of the 0.02 noise level were evaluated. Again, the three Post-ANOVA tests (DTK, HSD, and LSD) followed for a more detailed evaluation.

3.4.7.4 The ANOVA and Post-ANOVA Tests for Variable Type

With the highest performing color space identified, an ANOVA test was conducted to evaluate the three variable types: spatial location (XYZ), spatial relationship (MV), and the thematic attributes (highest performing color). The tests used similar data, information, and procedures as mentioned above for all variable sets: the Avg. AUC of 19 iterations of the 0.02 noise level were evaluated. Again, the three Post-ANOVA tests (DTK, HSD, and LSD) followed for a more detailed evaluation.

3.4.8 <u>3D Perspective Plots</u>

In addition to the quantitative metrics, it was important to see which particular points were grouped together. For a visual representation of the data, 3D perspective plots were generated from an oblique viewpoint, looking from the top, left, and front perspective. Spheres were used to symbolize each point to aid the perception of three-dimensional depth and understanding of which points were in the foreground and which were in the background. The points were colored based upon their membership in an output group. For most of the plots, if a point was not in a group it was appropriately invisible, though a few plots would show these as white. These plots offered a valuable way to view and interpret the patterns of grouped points. Selected plots are shown in the Results and Discussion section.

3.4.9 <u>K-means Clustering Comparison</u>

In contrast to the Mean Shift clustering, the K-means method had no bandwidths, so only one set of results was produced. Segmentation performance curves, AUC rankings, and segmentation surface plots were generated and 3D perspective plots were produced for visual analysis. Examples of the output are in the Results and Discussion section along with a comparison of the parameters of the two clustering methods.

3.4.10 LiDAR Case Studies

Building on the understanding of the Mean Shift clustering performance from the synthetic data experiments, my attention turned to understanding a real-world scene. The Indoor and Archaeological Site data sets were used. The same Mean Shift clustering methods were employed. The synthetic experimental results guided attention to particular bandwidths and variable sets.

To perform a quantitative analysis with the segmentation performance curves, Max. AUC and Avg. AUC rankings, and segmentation surface plots, it was necessary to select groups of points representing objects or features in the scenes. The Indoor LiDAR scan was labeled as 9 groups: 8 reference groups representing objects and one background group. The 8 reference objects were the following items: the brown box, red toolbox, blue pail, black case, green case, yellow ball, blue cup, and yellow ducks. For the Archaeological Site LiDAR data, the entire scan was labeled as three groups representing features in the landscape: the grass, the dirt, and the stone canal. Results were generated by employing the Mean Shift clustering to all of the initial 15 variable set combinations and 8 additional variable sets for alternative color spaces with the lightness removed. The quantitative accuracy assessment was performed and generated segmentation performance curves, the Max. AUC and Avg. AUC rankings, and segmentation surface plots. Finally, 3D perspective plots were also generated for a qualitative visual assessment and discussion of potential applications for geographic landscape representation.

Chapter 4 - RESULTS and DISCUSSION

4.1 Overview

The main goal of this thesis is to develop a concept and a means of finding features in a landscape. As there is a potential for abundant 3D data sets augmented with additional attribute information, it is beneficial to aggregate the observed data points into generalized regions of homogeneous characteristics. These regions may represent landscape features or components of more complex objects. Groups of data points having similar spatial locations, spatial relationships, and thematic attributes might be found with clustering techniques such as the Mean Shift method. First, the experiments with synthetic data provided a controlled environment and included pre-selected reference groups for the accuracy assessment of 3D point segmentation. Then, for the case studies, the methods were applied to real world data. Though the case studies also included an investigation into more complex variation, such as spatial and spectral variation (e.g., rugged surfaces and shadows), knowledge gained from the synthetic data analyses guided the interpretation of the case study results. There was an abundance of output from the experimental analysis; I present a condensed assessment of that output in this chapter.

4.2 Synthetic Data Set

The first result of my processing was simply the creation of an initial synthetic data set consisting of 3D (XYZ) point data with thematic (RGB) attributes, shown below.



Figure 12. Initial noise-free data set.

This initial noise-free data set served as the basis for the selection of ontological properties for an intuitive selection of reference groups. The planar regions of each side were assumed to have distinctive spatial locations (XYZ) and relationships (MV) compared to other sides. So, the points representing each side might be clustered together to identify aggregated groups of points representing each planar region. Though the sides were components of the more complex cube object, each side of the cube also had distinctive thematic attributes (RGB, LUV, or LAB), which helped to intuitively identify them as uniquely different features or reference groups.

4.3 K-means clustering

Using the noise-free data set, a brief comparison between the well-known K-means and the more recent Mean Shift clustering was done to describe differences in the processes and results. As an input parameter, the K-means requires the number of output groups to be specified. This is not commonly known for many data sets. However, 6 reference groups were labeled in the synthetic data set so 6 groups were specified for the K-means. Instead of specifying the number of output

groups, the Mean Shift clustering requires a bandwidth distance of evaluation. As mentioned in the Methods section, segmentation performance curves for 30 Mean Shift bandwidths were overlaid on the graphs. A comparison of selected segmentation performance curves is shown in Figure 13. The black lines show the number of correctly segmented groups found at each tolerance level. The K-means clustering yielded one set of results for each variable set. For comparing to the K-means, the highest performing of the 30 Mean Shift bandwidth results are shown in black (lower performing bandwidths are shown in gray).



Figure 13. K-means and Mean Shift Segmentation Performance Curves.

The segmentation performance curves showed that the Mean Shift clustering performs better than the K-means. One of the highest performing K-means variable sets, XYZ-MV-LAB, was matched by the Mean Shift with both meeting the ideal number of 6 reference groups up to a tolerance value of 1.0, or 100%. When the color variable set was switched, giving the XYZ-MV-RGB set, the K-means only reached three correctly segmented groups up to a tolerance value of 0.9, dropping down to two groups when evaluated against a tolerance value of 1.0. This means there were three output groups with each having at least 90% of the points of some corresponding reference group, though only two of the groups contained a full 100% of the points in the reference groups. The Mean Shift again had at least one bandwidth that found all of the 6 reference groups perfectly. With only two characteristics included, the Mean Shift was still able to maintain a high performance, while the K-means remained lower. However, when one of the two variable sets was the spatial relationship MV variable set, even the Mean Shift dropped from having 6 correctly segmented groups through the 0.9 tolerance value down to zero correctly segmented groups at the 1.0 tolerance. The K-means had even less success and only had two correctly segmented groups up to the 0.6 tolerance, meaning the two groups had at least 60% of their points in common with some reference groups. These examples show the Mean Shift clustering performing much better than the K-means, even when the K-means was set to find the exact number of reference groups in the noise-free data.

Used in computer science, computer vision, and machine learning disciplines, the conventional segmentation performance plots offer some level of quantitative and visual insight into the character of the resulting output segments. With many experiments to compare, however, the

first extension of the accuracy assessment was introduced, the area under the curve. Since the Mean Shift included 30 bandwidths for each variable set in each experiment, both the Max. AUC and the Avg. AUC were calculated, as described in the Methods section. The K-means AUC results and the Mean Shift Max. AUC results are shown in Table 12 as a ranked graph with the highest performing variable sets on top.

K-means	
AUC	Variable Set
1.00	RGB
1.00	LUV
1.00	LAB
1.00	XYZ-LAB
1.00	XYZ-MV-LAB
0.50	XYZ-RGB
0.50	XYZ-LUV
0.48	XYZ-MV-RGB
0.47	XYZ-MV-LUV
0.10	XYZ-MV
0.10	MV-LUV
0.03	MV-LAB
0.00	XYZ
0.00	MV
0.00	MV-RGB

Mean Shift	
Max AUC	Variable Set
1.00	XYZ
1.00	RGB
1.00	LUV
1.00	LAB
1.00	XYZ-RGB
1.00	XYZ-LUV
1.00	XYZ-LAB
1.00	XYZ-MV-RGB
1.00	XYZ-MV-LUV
1.00	XYZ-MV-LAB
0.90	XYZ-MV
0.02	MV-RGB
0.00	MV
0.00	MV-LUV
0.00	MV-LAB

Table 12. K-means and Mean Shift AUC rankings.

This table shows that, for many variable sets, some bandwidth in Mean Shift resulted in a higher Max. AUC than the AUC of the K-means method. In fact, with this noise-free data set, the Mean Shift had a perfect Max. AUC value of 1.0 for most of the variable sets and their combinations, meaning that through the tolerance value of 1.0, 100% of the points were correctly segmented into 6 groups matching the reference groups. The only drop off in Mean Shift performance was where the MV variable set was alone or in small combinations. The K-means seemed to do well

with some variable set combinations with color included, especially the LAB color space values. Overall, the common K-means clustering results rarely matched the performance of the highest performing Mean Shift bandwidths.

Affecting the results was a difference in the way each clustering process identifies clusters. With the K-means, the initial cluster seed locations are random and eventually move to the highest density of nearby points. With the Mean Shift, the bandwidth of evaluation is placed on every point, moving to the local center of density. The Mean Shift may be a more comprehensive search and offers the flexibility of finding any number of modes, or cluster centers.

For a visual evaluation, several 3D perspective plots from the K-means are shown in Figure 14; the Mean Shift plots are discussed in later sections.



Figure 14. K-means 3D perspective plots.

These plots offered an opportunity to learn more about the characteristics of the clustering results. Each variable set had particular properties and therefore particular tendencies in the clustering process. For instance, the XYZ variable set was intended to provide a measure of proximity, with similar valued points clustered. The expected output was locally contiguous groups of points. This was viewed as essential to separating different sides of the cube, or similar objects in a landscape separated by some distance. The 3D perspective plot for the XYZ variable set shows an unexpected K-means clustering result. The points from three sides, near corners of the cube, were grouped together. This is a reasonable outcome: points near a corner on one side are closer to points on a second side than they are to other points on the same side.

The 3D perspective plot for the MV variable set also revealed a partially unexpected result. The MV variable was intended to provide a value of each point's relationship to the neighboring points in an unstructured 3D space. As described in the Conceptual Framework, by averaging the vector values from a point to its neighbors, the Mean Vector was shown to have different values in a planar region than near an edge or in an area of high variability. The formula and values did follow that presumption, MV values of points representing planar regions had different values than those near edges. It was assumed that the XYZ variable set would help to separate planar regions with different locations or orientations. It was also expected that points near edges would have different values. However, it was not expected to majorly degrade planar regions near the edges. To notable effects arose. First, the values for all planar regions were so similar that they were all clustered together. Second, the number of points representing sides were almost always reduced (and effectively degraded) because the edges and corner points were separated into other

clusters. Though this may be helpful for applications interested in finding planar regions and edges, it was not perfectly matched for finding the solely planar regions. The 3D perspective plot for the 6 K-means groups clearly shows the edges as separated into different groups from the sides. Additionally, since only 6 groups were specified, the edges along the back side were not separated from the group of all planar regions. If 7 groups had been specified, the edges may have been separated from the 6 planar regions. Though the MV may be helpful for finding edges, for this thesis the MV generally degraded the output groups by reducing the number of points per side.

The 3D perspective plot for the K-means clustering of the RGB variable set shows a perfect segmentation. All 6 reference groups were perfectly segmented based upon the RGB values of 0, 0.5, 1 for red, green, and blue as specified in the Methods section. Though this was very helpful for clustering, it is important to note that this synthetic data structure greatly influenced the outcome. The small number of unique RGB values were very separable compared to the uniform distribution of more XYZ values (0-9) or the highly variable and less separable values of the MV, LUV, and LAB variable sets. Thus, the distribution of the RGB values and the structure of the initial synthetic data set appeared to boost the segmentation results of the variable sets including the RGB values.

The XYZ-RGB 3D perspective plot shows strong RGB influence. Recalling from the K-means segmentation performance curve and the AUC value, the XYZ-RGB variable set had three correctly segmented groups throughout all of the tolerance values. The 3D plot shows three sides perfectly grouped (left, front, and right), two sides as one group (top and bottom), and one side

split in half (back). The XYZ component may have influenced the splitting of the back side, though it did not create groups near corners as did the individual variable set. The RGB portion held together three sides individually and two sides as a group (the top and bottom were both levels of green, 1.0 and 0.5, respectively.) This begins to show that the synthetic experiments were strongly influenced by the pre-defined structure of the RGB variable set.

The K-means has been shown to have lower performance than the Mean Shift using several assessment aids. However, the K-means clustering was beneficial as it provided a baseline comparison of results, as well as insight into the processes and parameters of the two techniques. The visual evaluation helped to further understanding of the segmentation performance curves and the AUC rankings. It also offered an opportunity to discuss some of the key issues involving the synthetic data and XYZ, MV, and RGB variable sets. The remaining results and discussion concern the higher performing Mean Shift clustering method, as it shows relatively higher potential for being able to identify homogeneous regions in the landscape and group observations together by spatial location, spatial relationship, and thematic attributes.

4.4 Spatial Resolution

The results for the low-resolution (2 x 2 points per side) and the high resolution (44 x 44 points per side) data sets will now be presented and compared to the noise-free data set. Starting with the low resolution data, the selected segmentation performance curves and Max. AUC rankings, Figure 15 and Table 13, respectively, show that there was a maximum of one correctly segmented group throughout all of the variable sets and Mean Shift bandwidths. The MV again degraded the results, especially considering the 8 nearest neighbors for the calculation used points from 3 sides. This shows that if the MV variable were used to identify edges, the number

of neighbors and the relationship between distances to planar and non-planar neighbors is an important consideration.



Figure 15. Low resolution segmentation performance curves.

Max AUC	Variable set
0.17	XYZ
0.17	MV
0.17	RGB
0.17	XYZ-MV
0.17	XYZ-RGB
0.17	XYZ-LUV
0.17	XYZ-LAB
0.17	MV-RGB
0.17	MV-LUV
0.17	MV-LAB
0.17	XYZ-MV-RGB
0.17	XYZ-MV-LUV
0.17	XYZ-MV-LAB
0.00	LUV
0.00	LAB

Table 13. Low resolution Max. AUC rankings.

For the high resolution data set, the segmentation performance curves shown in Figure 16 are examples that the variable sets yielded between zero and 6 correctly segmented output groups. Apparently, the Mean Shift bandwidths reached certain distance thresholds that abruptly changed

the performance, but were stable through all tolerance values. The performance held steady through all tolerance values for every variable set, except the XYZ variable set.



Figure 16. High resolution segmentation performance curves.

This evaluation of resolution made use of multiple bandwidths for the Mean Shift clustering. Though the Max. AUC may have been achieved for at least one bandwidth, it was difficult to determine how many of the 30 overlaid bandwidths reached the maximum. For example, the XYZ, XYZ-LAB, and XYZ-RGB variable sets in Figure 16 may have all reached the maximum at some point, but it was more important to know the degree to which each variable set's highest performance persisted through all of the bandwidths evaluated. To address this aim, the Avg. AUC (averaged across all bandwidths) for each variable set was calculated. This quantitative metric provided a number with which to rank a variable sets' overall performance across all of the bandwidths evaluated. In summary, the Max. AUC can show which variable sets had the highest performing bandwidths and the Avg. AUC can show which variable sets performed well over all of the bandwidths. The complementary high resolution Max. AUC and Avg. AUC rankings are shown in Table 14.

Max AUC	Variable Set	Avg AUC	Variable Set
1.00	XYZ	0.30	XYZ-RGB
1.00	RGB	0.27	XYZ-LUV
1.00	LUV	0.27	XYZ-LAB
1.00	LAB	0.20	RGB
1.00	XYZ-RGB	0.20	LUV
1.00	XYZ-LUV	0.19	LAB
1.00	XYZ-LAB	0.19	XYZ
0.00	MV	0.00	MV
0.00	XYZ-MV	0.00	XYZ-MV
0.00	MV-RGB	0.00	MV-RGB
0.00	MV-LUV	0.00	MV-LUV
0.00	MV-LAB	0.00	MV-LAB
0.00	XYZ-MV-RGB	0.00	XYZ-MV-RGB
0.00	XYZ-MV-LUV	0.00	XYZ-MV-LUV
0.00	XYZ-MV-LAB	0.00	XYZ-MV-LAB

Table 14: High resolution Max. AUC and Avg. AUC rankings.

Beyond the Max. AUC and Avg. AUC rankings, a second innovation uses the segmentation performance curve information to create new segmentation surface plots which illustrate the performance over every bandwidth. These plots provided an additional means of evaluating the quantitative measures of segmentation performance over all of the bandwidths using a single visualization. As described in the Methods section, the x-axis represents the tolerance values (0.5 to 1.0), the y-axis represents each bandwidth (0.0 to 2.9 standard deviation units), and all locations in the graph contain color patches indicating the number of correctly segmented groups



(0 to 6, from white to black, respectively.) Examples are shown Figure 17.

Figure 17. High resolution segmentation surface plots.

The new segmentation surface plots readily display the number of correctly segmented output groups in shades of gray from zero (white) to 6 (black). The XYZ-RGB variable set, with an Avg. AUC of 0.30, has many bandwidths that have the ideal 6 groups correctly segmented in bandwidths ranging from 0.3 to 1.1 standard deviation units. The XYZ-LAB had an equal Max. AUC of 6 groups, but a slightly lower Avg. AUC of 0.27. This was apparently due to fewer correctly segmented groups (in levels of gray) in the 1.0 and 1.1 bandwidths; lower bandwidths to 0.3 performed as well as the XYZ-RGB. The segmentation surface plot for the XYZ set shows that its performance peaked at 6 groups, but dropped off at varying tolerance levels for some bandwidths. This was a curious effect, that performance would increase, then decrease, and increase again before finally degrading with higher bandwidths. To investigate which points were being grouped together for the high resolution data set, the 3D perspective plots are shown in Figure 18.



Figure 18. High resolution 3D perspective plots.

The 3D perspective plots revealed several unexpected results for the XYZ variable set. The 0.1 bandwidth (whose plot is a closer view and not at the same scale as the others) was apparently the lower threshold for generating a corner grouping as previously noted for the XYZ variable set However, only one corner was created; why other corners were not found was unclear. At the 0.3 bandwidth, rectangular regions on each side were grouped, though it was still not enough points to meet the majority criteria for correctly segmented groups. At the 0.5 bandwidth, the circular edges reappeared in the spatial pattern, at least indicating that some points were being clustered at some distance from the center of each side. At bandwidth 0.6, the high resolution XYZ variable set clustered 6 correctly segmented groups in a circular pattern from the center of each side; this persisted through the 0.8 tolerance value as shown in the segmentation surface plot. At a larger bandwidth of 0.7, the circular shape reduced in size and the number of clustered points decreased, with the 6 correctly segmented groups having over 70% of the points in the reference group, a drop from the previous bandwidth's percentage of over 80%. The only

explanation I can provide for this is the possibility that the bandwidth is now just large enough to create competition between different modes that the Mean Shift cannot resolve. As the bandwidth grows to 0.9, the circle grows again and maintains 6 correctly segmented groups through a tolerance value of 0.8, or over 80% of the reference points. Eventually, between the 1.0 and 1.2 standard deviation bandwidths, the variable set reaches a perfectly segmented cube; each planar side has 100% of the points included in an output group. This is shown in the both the 3D perspective plot (Figure 18) and the segmentation surface plot (Figure 17). Finally, at the bandwidths of 1.3 and higher, the points are contained in one single group. This undersegmentation is a common result for large bandwidths in all variable sets. This evaluation demonstrates the complementary value of the new segmentation surface plots and the 3D perspective plots.

To compare the different resolutions (point spacings of 0.2, 1.0, and 3.0), the visualizations offered great potential for revealing the character of the clustering results. However, to make a final determination of the importance of various levels of resolution, I made use of the Avg. AUC metrics, shown in Table 15. It was a convenient way to quantitatively rank the variable sets and reach conclusions about the different resolutions. The rankings show that the lower resolution was problematic and there were too few points to find locally homogeneous groups. The higher resolution provided numerous points and intriguing results. However, it did not appear necessary to have a very high density of data to correctly segment the points. The moderate spacing of 1.0 units actually had an equal or higher Avg. AUC for most variable sets; it performed much better with the XYZ variable set and even had some success with the MV variable set.

Spacing 3.0		Spacing 1.0			Spacing 0.2		
Avg AUC	Variable set	Avg AUC Variable set			Avg AUC	Variable set	
0.06	XYZ	0.31	XYZ-RGB		0.30	XYZ-RGB	
0.06	MV	0.28	XYZ-LUV		0.27	XYZ-LUV	
0.06	XYZ-RGB	0.28	XYZ		0.27	XYZ-LAB	
0.06	MV-RGB	0.27	XYZ-LAB		0.20	RGB	
0.06	MV-LUV	0.20	RGB		0.20	LUV	
0.06	MV-LAB	0.20	LUV		0.19	LAB	
0.06	XYZ-LUV	0.19	LAB		0.19	XYZ	
0.06	XYZ-LAB	0.15	XYZ-MV-LUV		0.00	MV	
0.05	XYZ-MV-RGB	0.13	XYZ-MV-RGB		0.00	XYZ-MV	
0.05	XYZ-MV-LUV	0.13	XYZ-MV-LAB		0.00	MV-RGB	
0.04	XYZ-MV-LAB	0.03	XYZ-MV		0.00	MV-LUV	
0.04	XYZ-MV	0.00	MV-RGB		0.00	MV-LAB	
0.03	RGB	0.00	MV		0.00	XYZ-MV-RGB	
0.00	LUV	0.00 MV-LUV			0.00	XYZ-MV-LUV	
0.00	LAB	0.00	MV-LAB		0.00	XYZ-MV-LAB	

 Table 15. Comparison of various resolutions' Avg. AUC.

4.5 Noise Evaluation

The experiments up to this point only made use of noise-free data sets; equally spaced spatial patterns and perfectly selected attribute information. In practice, any investigation is subject to error. The source of error may be in the experimental design, data collection, processing steps, or analyses. Considering 3D spatial point data, observational equipment (e.g., range scanners, GPS receivers) or complex environmental variability (e.g., leaves on vegetation, varying temperature or moisture) could potentially lead to noisy data sets with some level of uncertainty as to the location and boundaries of landscape features. Before working with real world data, I created 19 sets of data for 5 levels of noise in order to gain a more thorough understanding of the variable sets and Mean Shift clustering results. The synthetic experiments leveraged computational tools,

generated data samples with increasing levels of noise, and produced the most reliable results of this thesis through statistical analyses and hypothesis testing.

The Methods section provided details of the data set production. As a visual reference, the examples below show in Figure 19 the spatial variability of the noise-free data and sets with 5 levels of increasing noise. The (non-noisy) colors represent the 6 reference groups (the plots are at different scales to show outliers.)



Figure 19. Examples of adding noise to the spatial data.

Only one noise-free data set was created and 19 data sets of each level of noise were created. The noise, randomly selected from a Gaussian distribution having a mean of zero and variance ranging from 0.02 to 0.10 in increments of 0.02 was added to the spatial (XYZ) and thematic (RGB) values. Next, other derived values (MV, LUV, LAB) were processed and all variable sets were standardized as with the noise-free data set. The 19 sets of noisy data provided a sufficient number of random samples to perform statistical analyses. The accuracy assessment was performed on all the data sets in this experiment to find the number of correctly segmented groups at each bandwidth for each variable set. Then, the Avg. AUC values were calculated as measures of overall performance of each variable set; this value was used for all statistical analyses.

4.5.1 <u>One-sample t-tests for All Variable Sets at All Levels of Noise</u>

The first aim was to determine what level of noise produced significantly lower results compared to the noise-free data set. A one-tailed, one-sample t-test was performed for each variable set and each level of noise, comparing the mean of the 19 samples of each level of noise to the noise-free value. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The hypotheses are presented below and the Mean Avg. AUC and p-values are shown in Table 16.

- Ho: the mean of the 19 noisy Avg. AUC results for each variable set = the noise-free Avg. AUC for each variable set
- Ha: the mean of the 19 noisy Avg. AUC results for each variable set \leq the noise-free Avg. AUC for each variable set

		Noise-free	With-noise									
		0.00	0.02		0.04		0.06		0.08		0.10	
XYZ	Mean Avg. AUC	0.275	0.243		0.174		0.101		0.061		0.041	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
MV	Mean Avg. AUC	0.000	0.000		0.000		0.000		0.000		0.000	
	P-value		-		-		-		-		-	
RGB	Mean Avg. AUC	0.200	0.206		0.198		0.182		0.165		0.153	
	P-value		-		0.065		<<0.001	***	<<0.001	***	<<0.001	***
LUV	Mean Avg. AUC	0.200	0.134		0.083		0.045		0.030		0.018	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
LAB	Mean Avg. AUC	0.189	0.122		0.077		0.043		0.026		0.017	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
XYZ-MV	Mean Avg. AUC	0.032	0.011		0.009		0.006		0.005		0.005	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
XYZ-RGB	Mean Avg. AUC	0.306	0.304		0.295		0.281		0.264		0.253	
	P-value		0.159		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
XYZ-LUV	Mean Avg. AUC	0.283	0.266		0.228		0.171		0.125		0.092	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
XYZ-LAB	Mean Avg. AUC	0.272	0.260		0.217		0.186		0.149		0.128	
	P-value		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
MV-RGB	Mean Avg. AUC	0.002	0.008		0.031		0.030		0.019		0.014	
	P-value		0.999		0.999		0.999		0.999		0.999	
MV-LUV	Mean Avg. AUC	0.000	0.007		0.008		0.004		0.002		0.001	
	P-value		0.999		0.999		0.999		0.999		0.977	
MV-LAB	Mean Avg. AUC	0.000	0.004		0.008		0.004		0.002		0.001	
	P-value		0.999		0.999		0.999		0.999		0.999	
XYZ-MV-RGB	Mean Avg. AUC	0.128	0.168		0.164		0.155		0.138		0.126	
	P-value		0.999		0.999		0.999		0.999		0.133	
XYZ-MV-LUV	Mean Avg. AUC	0.145	0.154		0.121		0.084		0.050		0.030	
	P-value		0.999		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***
XYZ-MV-LAB	Mean Avg. AUC	0.128	0.137		0.115		0.093		0.068		0.050	
	P-value		0.998		<<0.001	***	<<0.001	***	<<0.001	***	<<0.001	***

*** indicates significance at $\alpha = 0.001$

Table 16. Mean Avg. AUC and P-values for one-tailed, one-sample t-tests.

Based on the fact that many noise-free variable sets produced perfect segmentations (6 correctlysegmented groups through tolerance levels ranging from 0.5 to 1.0), I expected a small amount of noise would not significantly degrade the performance. However, the p-values from the onesample t-test showed that even a small amount of noise produced significantly lower results than the noise-free data. Because the noisy samples for each variable set and level of noise had small variability, they produced very small margins of error (~ ± 0.003). This provided assurance that the true mean value for each variable set and level of noise was most likely within small 95% confidence intervals around the sampled means.

There were a few issues with the tests that require explanation. The 19 RGB data sets with 0.02 noise yielded the same exact Avg. AUC value. There was no variation in the samples and the t-test resulted in a null value. In similar fashion, the MV variable set performed poorly in the noise-free data and could not perform any worse with noise. The zero values generated a null output from the t-test. Most of the other combined variable sets which included the MV set actually managed to improve the Avg. AUC slightly for low levels of noise. This was probably due to the noise actually reducing the negative effect of the Mean Vector, allowing the other variables (e.g., XYZ, RGB, etc.) to have a positive effect on the performance until high levels of noise eventually degraded the entire set. Regardless of any slight improvement, most of the t-tests with the MV set did not have significantly lower values since they were very low in the noise-free set.

Nearly all of the other noisy variable sets performed significantly lower than the noise-free values. The only variable sets that were not significantly lower were the XYZ-RGB at the 0.02 noise level and the RGB at the 0.04 noise level. Both of these variable sets performed very well compared to the noise-free sets and likely benefitted from the strong separability of the RGB values, as previously noted.

4.5.2 <u>Selection of the 0.02 Noise Level for Additional Analyses</u>

One of the purposes of evaluating which level of noise was not significantly different from the noise-free level was to provide a means of obtaining many samples. In practice, a two percent

level of noise could be considered common, whether due to variability in the environment, or errors in observations or sensors. Though the 0.02 noise level was often significantly lower for many of the variable sets, the 0.02 level of noise was selected as a reasonably low level to generate enough randomly sampled synthetic realizations for further statistical analyses. This was deemed acceptable because the values of the highly performing RGB variable set indicated a performance threshold near the 0.02 level of noise. The RGB values had no variance at the 0.02 level of noise; reducing the level of noise may result in more variable sets with no variance. Increasing the level of noise resulted in more variable sets having significantly lower performance.

4.6 Comparison of Multiple Means with Analysis of Variance (ANOVA) tests

The objectives of this thesis were to determine the highest performing variable sets, color spaces, and which type of variable set was most helpful (spatial location, spatial relationship, or thematic attribute.) It was therefore necessary to compare multiple mean values. The One-way ANOVA test is suitable for comparing multiple means with different variances and sample sizes. This test determines if the means are equal or if any one of them is significantly different. A Post-ANOVA test is necessary to determine which particular mean values are significantly different. The three Post-ANOVA tests I employed were the Dunnett-Tukey-Kramer (DTK) test, the Tukey Honestly Significant Differences (HSD) test, and the Least Significant Differences (LSD). They are all very closely related and generally find the least significant differences for each mean. The DTK and HSD tests result in sets of pairwise comparisons. These two tests had similar, though not exactly the same, results. The differences are attributable to the numerous adjustments available as extensions to the main test. The LSD had similar results regarding the mean, confidence intervals, and the least significant difference values. However, this test simply

outputs all of the means, the least significant differences and determines groups of means that are not significantly different. Since the LSD output is more clear than the multitude of pairwise comparisons, I first show the ANOVA results and the LSD results follow.

4.6.1 <u>The ANOVA and Post-ANOVA Tests for All Variable Sets</u>

Continuing investigation as to which variable sets have the highest performance, all of the Avg. AUC values from 19 samples (from the 0.02 noise level) of 15 variable sets were used for the ANOVA test. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The hypotheses are presented below and the results are shown in Table 17. The p-value shows there is a significantly low probability that the means are equal; as such, the null hypothesis is rejected.

Ho: the mean of the 19 noisy Avg. AUC results for variable set (1)
= the mean of the 19 noisy Avg. AUC results for variable set (2)
= ...

= the mean of the 19 noisy Avg. AUC results for variable set (15)

Ha: not all of the means of the 19 noisy Avg. AUC results for the variables sets were equal

Analysis of Variance Table: All Variable Sets									
Response: Avg. AUC									
	Df Sum Sq Mean Sq F value Pr(>F)								
Variable_Set 14 3.09301 0.220929 3889.4 < 2.2e-16 ***									
Residuals	270	0.01534	0.000057						

Table 17. ANOVA results for all variable sets.

The Least significant differences Post-ANOVA test followed. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The resulting graph, Figure 20, shows a bar for the mean of all 19

samples of each variable set and whiskers for the least significant difference for the mean of each variable set. Groups of means are also labeled with letters. Mean values belonging to the same group are not significantly different from each other.



Figure 20. Least significant differences results for all variable sets.

Compared to the one-sample, noise-free Avg. AUC ranking, this graph contains a more reliable ranking of variable sets because it was based on the mean and variance of 19 samples of a realistic value of 2% noise (0.02 variance). As noted previously, the margin of error and 95%

confidence intervals were sufficiently small, as well. The whiskers represent the least significant difference for each variable set's mean. The letters indicate groups of means that were not significantly different from each other; variable sets with a shared letter were not significantly different.

The most obvious pattern in the graph is that the MV and two-set combinations with MV performed very poorly. The performance of the MV improved only when used in a three-variable set combination. However, these are still relatively low compared to the XYZ and color variable sets. The MV evidently degrades segmentation performance. This indicates a conceptual mismatch between the selected reference groups and the implementation of the spatial relationship metric, MV. This metric resulted in different values (and thus different output groups) for points near some edges as compared to other edges, corners, and planar regions. This may be helpful for some applications. Unfortunately, for this experimental framework, removing the edge points from the planar points degraded the percentage of points belonging to groups representing sides.

The remaining results for the XYZ and color variable sets also offered an answer for the general hypotheses of this thesis: that multiple variable sets enhance the ability to find groups of data as compared to single variable sets. With the three-variable set combinations degraded by the MV, it was readily apparent that the two-variable set combinations of XYZ and color had significantly higher performance than the individual XYZ and color variable sets. Even the poorly performing MV variable set had boosted performance when combined with two additional variable sets. So, the identification of points representing regions of homogeneity is boosted by the use of

additional variable sets. Spatial location and thematic attributes help to increase the number of points correctly segmented as sides, or components, of the more complex 3D cube object. The least significant differences graph of all variable sets provided clear and encouraging results of higher performance with additional variable sets.

4.6.2 <u>The ANOVA and Post-ANOVA Tests for Color Variable Sets</u>

To identify the best performing color space, all of the Avg. AUC values from 19 samples (from the 0.02 noise level) of the 3 individual color space variable sets were used for the ANOVA test. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The hypotheses are presented below and results are shown in Table 18. The p-value shows there is a significantly low probability that the means are equal; as such, the null hypothesis is rejected.

- Ho: the mean of 19 noisy Avg. AUC results for the RGB variable set
 = the mean of 19 noisy Avg. AUC results for the LUV variable set
 = the mean of 19 noisy Avg. AUC results for the LAB variable set
- Ha: not all of the means of the 19 noisy Avg. AUC results for the RGB, LUV, and LAB variables sets were equal

Analysis of Variance Table: Color Variable Sets										
Response: Avg. AUC										
	Df Sum Sq Mean Sq F value Pr(>F)									
Variable_Set 2 0.07741 0.038704 817.54 < 2.2e-16 ***										
Residuals	54	0.00256	0.000047							

Table 18. ANOVA results for color variable sets.

The least significant differences Post-ANOVA test followed. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The resulting graph in Figure 21 shows a bar for the mean of all 19

samples of each color variable set and whiskers for the least significant difference for each group mean. Groups of means are also labeled with letters. Mean values belonging to the same letter group are not significantly different from each other.



Figure 21. Least significant differences results for color variable sets.

First, it is important to note that the RGB variable set did not have least significant distances because the 0.02 noise samples for RGB all had the same Avg. AUC value; there was no variance. The 19 RGB samples, therefore, had no margin of error. As the test results show, the RGB variable set had a significantly higher performance than the LUV, and LAB variable sets.

This was most likely related to the construction of the synthetic data set. The RGB values were much more separable and able to cluster much better than the LUV or LAB sets. For example, the color space and standardized values for both full intensity red and blue are shown in Table 19. The range and variance for the difference of the standardized RGB, LUV, and LAB values for red and blue are also shown. In the example, the RGB values had a much larger range and variance providing for highly separable clusters compared to the LUV and LAB values. The design of the synthetic data set provided values that helped the RGB variable set cluster more successfully. To overcome this, a wider variety of different colors should have been used. This would also more readily parallel the potential variation in thematic attributes in a real world investigation. Though the basic data used here were meant to be a simplistic representation in a test environment, it brought about useful insights relating to the clustering process, color variable sets, and thematic data values.

Variable Set	Color	Space Va	alues	Stand	ardized V	Range	Variance	
RGB (red)	1.00	0.00	0.00	1.96	-0.65	-0.65		
RGB (blue)	0.00	0.00	1.00	-0.65	-0.65	1.96		
Difference	1.00	0.00	-1.00	2.62	0.00	-2.62	5.23	6.84
							-	
LUV (red)	53.24	175.01	37.76	0.14	1.58	0.37		
LUV (blue)	32.30	-9.40	-130.34	-0.81	-0.34	-1.56		
Difference	20.94	184.42	168.09	0.95	1.93	1.93	0.98	0.32
LAB (red)	53.24	80.09	67.20	0.14	0.82	0.70		
LAB (blue)	32.30	79.19	-107.86	-0.81	0.81	-1.54		
Difference	20.94	0.91	175.06	0.95	0.01	2.24	2.22	1.25

Table 19. Color space value examples favoring clustering of the RGB variable set.
4.6.3 The ANOVA and Post-ANOVA Tests for Variable Types

Determining the highest performing variable type is also a research question in this thesis. Spatial location, spatial relationship, and thematic attributes are represented in this test by XYZ, MV, and RGB, respectively. RGB was chosen here because it was the highest performing color space, even though the data were created in a fashion that increased the possibility of RGB clustering successfully. Also, at this point in the evaluation, the MV variable set was expected to reduce the number of correctly segmented groups, though these statistical tests offer a confirmatory test. All of the Avg. AUC values from 19 samples (from the 0.02 noise level) of the 3 selected variable types were used for the ANOVA test. A rejection threshold alpha value of α = 0.05 was employed. The hypotheses are presented below and the results are shown in Table 20. The p-value shows there is a significantly low probability that the means are equal; as such, the null hypothesis is rejected.

- Ho: the mean of 19 noisy Avg. AUC results for the XYZ variable set
 = the mean of 19 noisy Avg. AUC results for the MV variable set
 = the mean of 19 noisy Avg. AUC results for the RGB variable set
- Ha: not all of the means of the 19 noisy Avg. AUC results for the XYZ, MV, and RGB variables sets were equal

Analysis of Variance Table: Variable Types							
Response: Avg. AUC							
	Df	Sum Sq	Mean Sq	F value	Pr(>F)		
Variable_Set	5	1.9979	0.39958	45.006	< 2.2e-16 ***		
Residuals	393	3.4892	0.00888				

Table 20. ANOVA results for variable types.

The Least significant differences Post-ANOVA test followed. A rejection threshold alpha value of $\alpha = 0.05$ was employed. The resulting graph, Figure 22, shows a bar for the mean of all 19

samples of each variable type and whiskers for the least significant difference of the group mean. Groups of means are also labeled with letters. Mean values belonging to the same letter group are not significantly different from each other.



Figure 22. Least significant differences results for variable types.

After looking at the previous tests of all variable sets and selecting the highest performing color variable set, RGB, this test was aimed at finding the most beneficial type of variable set: spatial location (XYZ), spatial relationship (MV), or thematic attribute (RGB). Already noting the

effects of MV, it was easy to deduce that performance was much higher without the MV and much lower with the MV. As for the XYZ and RGB variable sets, a higher mean was found with the XYZ, though it was not significantly higher than with the RGB. This is indicated by both "With XYZ" and "With RGB" belonging to the same group of means, "B". These two collections performed significantly better than without either of the XYZ or RGB sets, which were in a group of lower means, "C". This indicates that using the spatial and thematic variable sets improved segmentation performance in the experiments. The process of grouping points representing sides, or parts of the whole cube, benefitted from having both the XYZ and RGB data included.

4.7 Alternative Color Configuration

The alternative color configuration was intended to address the question of identifying regions having similar spatial location and relationship and differing in the values of thematic attributes. The previous noise evaluation did alter colors as well as spatial properties, though it was meant to impose increasing levels of variability in the data. The alternative color configuration used the noise-free spatial values and only altered the colors per side. Instead of one color per each side, different colors were applied to each half of a side. The initial red, green, and blue color combinations were used along with new combinations as noted in the Methods section. Adjacent sides, across edges of the cube, had either the same or different colors to test the clustering ability when encountering edges. The accuracy assessment tested the results for 12 reference groups containing 32 points per group. The segmentation performance curves only showed a few instances when the clustering produced one correctly segmented group which lasted through a tolerance value of 0.7, at the most. Two examples are shown in Figure 23 with the only two resulting patterns.



Figure 23. Alternative color configuration segmentation performance curves.

The variable set combination of XYZ and each color had the best results as shown on the left. Adding the MV to the XYZ and color variable sets again degraded the performance. Interestingly, none of the color variable sets correctly segmented groups individually. This suggested that when new color combinations were used (beyond solely red, green, or blue), the color variable sets did not have any success. The Max. AUC and Avg. AUC rankings, in Table 21 below, show the level of performance for the variable sets in the alternative color configuration.

Max AUC	12 Reference Groups (Alt. Color)	Avg AUC	12 Reference Groups (Alt. Color)
0.08	XYZ-RGB	0.01	XYZ-RGB
0.08	XYZ-LUV	0.01	XYZ-LUV
0.08	XYZ-LAB	0.01	XYZ-LAB
0.02	XYZ-MV	0.00	XYZ-MV-RGB
0.02	XYZ-MV-RGB	0.00	XYZ-MV-LUV
0.02	XYZ-MV-LUV	0.00	XYZ-MV-LAB
0.02	XYZ-MV-LAB	0.00	XYZ-MV
0.00	XYZ	0.00	XYZ
0.00	MV	0.00	MV
0.00	RGB	0.00	RGB
0.00	LUV	0.00	LUV
0.00	LAB	0.00	LAB
0.00	MV-RGB	0.00	MV-RGB
0.00	MV-LUV	0.00	MV-LUV
0.00	MV-LAB	0.00	MV-LAB

Table 21. Alternative color Max. AUC and Avg. AUC.

The 3D perspective plots offer the best insight into the clustering performance with the alternative color configuration, shown in Figure 24.



Figure 24. Alternative color configuration 3D perspective plots.

One of the highest performing variable sets, with only one correctly segmented group, was the XYZ-RGB. At a very small bandwidth of 0.1 standard deviation units, only 18 out of 32 points in the reference group were clustered together. At the 0.3 bandwidth, the group combined with an adjacent group across an edge of the cube. The MV-RGB variable set with a small 0.1 bandwidth grouped portions of several halves together into two groups. The points near edges were excluded, as was also the case for the individual MV variable set. The XYZ variable set managed to group three sides into separate groups and three other sides into one group. However, since the aim here was to find 12 reference groups, two groups per side, this result was not correctly segmented. The influence of thematic information was readily apparent with the RGB variable set. Halves of sides were in separate groups, since they had different colors. However, some halves on different sides were grouped together and not correctly segmented per the assumption of 12 reference groups. The individual XYZ and RGB variable sets exhibited some expected tendencies. The XYZ did group many points that were close to each other and the RGB and other color spaces grouped similar colors together. Though they clustered well per the input data values, they did not find the 12 separate reference groups. This effect supported the idea that features with different spatial or thematic attributes could result in correctly segmented groups with respect to the input data. However, the combinations of spatial and thematic variable sets did not perform as well. One additional observation, however, is that smaller bandwidths appeared to work better in the alternative color configuration. This may be due to the smaller distances between values of both the spatial and thematic attributes as compared to the other experiments with 6 groups. Bandwidths may be sensitive to the range of the variable set values, as well as the number and spacing of modes within each variable set (e.g., more sets of colors would have more modes for the thematic variable).

4.8 LiDAR Case Studies

The final investigations of this thesis involve data collected from real-world scenes. Terrestrial LiDAR scans were collected in two scenes, an Indoor and an Archaeological Site scene. The details of both are described in the previous Methods section. All of the same 15 variable set combinations were generated and supplemented with 8 new combinations including two new variable sets UV and AB (simply the LUV, and LAB without the lightness component). Since shadows were a concern, using color spaces that ignored lightness was expected to allow clustering features based upon their chromatic properties without the influence of lightness or shadows. As this was an extension from the synthetic data experiments, the same Mean Shift clustering methods, number of neighbors, and bandwidths were employed and the same accuracy assessment was performed.

4.8.1 Indoor Scan

The first case study, in a progression from the synthetic data, I initially worked with the Indoor LiDAR scan which contained spatial and thematic (spectral) data collected from simple constructed shapes having contrasting colors. Various objects in the scene had different properties. Some were planar, some were cylindrical, and some were spherical. The objects had different smoothness and colors. One notable challenge was to reduce the effect that shadows and lighting had on the segmentation. Though shadows may be helpful for some applications, it was considered beneficial to remove their effects to improve the correct segmentation of objects in this scene.

First, the Indoor LiDAR scan was sampled to boost computational efficiency. Then, the scan was labeled as 9 groups: 8 reference groups representing objects and one background group. The 8

reference objects were the brown box, the red toolbox, the blue pail, the black case, the green case, the yellow ball, the blue cup, and the yellow ducks. Figure 25 shows images of the unsampled data, the sampled data in RGB color, and the sampled data colored per their reference groups.



Figure 25. Indoor LiDAR scan, sampled data, and labeled reference groups.

After the Mean Shift clustering was complete, the accuracy assessment generated segmentation performance curves. Two of these are shown in Figure 26 and are examples of higher performing variable sets. The two images both show that 6 of the 9 reference groups were correctly segmented through the 0.6 tolerance level, meaning that the 6 groups had between 60-70% of their respective reference groups' points. Four output groups had over 80% and three output groups had over 90% of their respective reference groups.



Figure 26. Indoor LiDAR segmentation performance curve examples.

The Max. AUC and Avg. AUC rankings, shown in Table 22, allow for a performance evaluation of all variable sets in relation to each other. The Max. AUC, relating directly to the largest area under the curve for each variable, shows that the alternative color spaces generally produce the highest performing results, whether they include the lightness variable or not. The XYZ and MV variable set are spread throughout the ranking list, however the XYZ set is evidently associated with some of the highest Max. AUC values. On the other hand, the RGB variable set is included in the lowest of the Max. AUC values. When all of the bandwidths are averaged for the Avg. AUC, a relative measure of performance is generated. The ranking of the Avg. AUC values reinforces the fact that the UV, AB, LAB, and LUV variable sets produce higher levels of correct segmentation overall. The XYZ variable set follows and generally performs better than the MV or RGB variable sets.

Max AUC	Variable Set	Avg AUC	Variable Set
0.47	LAB	0.28	UV
0.47	XYZ-LUV	0.28	AB
0.47	XYZ-LAB	0.27	LAB
0.47	XYZ-MV-LUV	0.26	LUV
0.47	XYZ-MV-LAB	0.24	XYZ-LUV
0.42	LUV	0.23	XYZ-LAB
0.41	AB	0.21	XYZ-UV
0.39	UV	0.20	XYZ-AB
0.39	XYZ-AB	0.15	XYZ-MV-LUV
0.35	MV-LUV	0.14	XYZ-MV-LAB
0.33	XYZ-UV	0.13	MV-LUV
0.33	MV-LAB	0.11	MV-LAB
0.33	XYZ-MV-AB	0.11	MV-UV
0.31	MV-AB	0.11	XYZ-MV-UV
0.31	XYZ-MV-UV	0.11	XYZ-MV-AB
0.29	XYZ-MV-RGB	0.09	MV-AB
0.28	XYZ-RGB	0.09	XYZ-RGB
0.28	MV-UV	0.07	RGB
0.27	RGB	0.06	XYZ-MV-RGB
0.10	XYZ-MV	0.03	XYZ
0.09	XYZ	0.03	MV-RGB
0.08	MV-RGB	0.01	XYZ-MV
0.00	MV	0.00	MV

Table 22. Indoor LiDAR Max. AUC and Avg. AUC values.

The segmentation surface plots allow for a quick evaluation of the performance of all of the variable sets throughout all of their Mean Shift bandwidths. Guided in part by the AUC rankings, a selection of 6 segmentation surface plots are shown in Figure 27. The UV and AB variable sets plots are similar and illustrate their ability to segment throughout a large range of bandwidths. Though performing well overall, their peak performance is not as high as the LAB or the XYZ-LAB and XYZ-LUV combinations. Though these three sets are again similar and show that the addition of XYZ slightly degrades the performance at small and large bandwidths. This is possibly due to the different real world objects sizes, and therefore different XYZ distances in

the data. The lower performing XYZ-RGB variable set is also shown to illustrate the difference between the synthetic data and the real world data. The highly separable distribution of RGB values in the synthetic data readily generated clusters and the variable set had much higher performance compared to the LUV and LAB color spaces. However, in the real world data, the RGB values were less separable. The alternative color spaces evidently had values that were much more separable, allowing for improved clustering results that may be in line with the way different colors in the landscape are perceived by humans.



Figure 27. Indoor LiDAR segmentation surface plot examples.

The evaluation also involved a subjective visual inspection regarding particular objects represented in the source data set and the properties of their variables. Four Mean Shift clustering results are shown in Figure 28 as 3D perspective plots for discussion purposes. The plots for the MV variable sets were not selected for discussion here because they separated points representing edges from the output groups.



Figure 28. Indoor LiDAR 3D perspective plot examples.

Smaller Mean Shift bandwidths (~ 0.3) segmented the Indoor scene better than the preferred bandwidths in the synthetic data (~ 0.6). This reinforces the notion that the bandwidths may need to be selected to fit the distribution of data values and size of the clusters. In fact, the only three objects that rarely clustered were the small objects with very few points (the yellow ball, the blue cup, and the yellow ducks) as they had an entirely different scale and were relatively undersampled in the scan. The XYZ-LUV results illustrate the effect of the spatial location variables. Larger regions that exhibited similar thematic (color) characteristics were segmented into different groups by the XYZ variable set. This XYZ-LUV set was also sensitive to lightness; regions with different lighting or shading were in separate groups, as seen in the upper right region of the bookshelves. The XYZ-LAB variable set continued to break up similarly colored features, such as the brown wood box at the bottom of the scene. Both of these variable sets created many groups for the objects on the shelves. Over-segmentation could be viewed as preferable in this application since it is easier to combine groups of data than it is to separate groups. The LAB variable set, on the other hand, had less over-segmentation, yet effectively generalized the scene. Some of the features, however, were not found at all (the two yellow ducks in the center of the shelves, and the yellow ball) and some were combined with larger groups (the blue coffee cup toward the upper left.) With this set, the shadow region in the upper right was not found, possibly because the shadow was not as influential as in other regions. Of particular interest, is the black rectangular case on the left side of the book shelf. It was easily segmented as a separate group with the LAB variable set. However, when the lightness variable was removed for the AB variable set, the black case was combined with the white background. Though the AB variable set did well at removing the effect of shadows, the clustering process could not find any difference between black case and white background points for two reasons: lightness was not used for clustering and both of these extremes in lightness (black and white) do not have much difference in color. In practice, it may be helpful to remove the effect of shadows, though one should exercise caution if clustering white and black objects is not desirable. To summarize, the XYZ variable set separated the data into local regions, the MV separated edges from planar regions, and the thematic (color) attributes did the best at segmenting easily recognizable objects in the scene. The inclusion or exclusion of lightness is a potential option for refinement of spectral segmentation of a real world scene.

4.8.2 <u>Archaeological Site Scan</u>

The second case study was the Archaeological Site LiDAR scan. The scene, shown in Figure 29 contained nearly flat (not necessarily level) soil in the right foreground, a stone-lined and capped water canal running diagonally from the lower left foreground to the upper right of the scene, and grassy vegetation toward the left background portion. This scan provided rugged features with the stones in the water canal which also had highly variable color values. The grassy

vegetation and soil provided relatively flat surfaces, though the blades of grass presented some variation. Intuitively, the spectral values help to quickly differentiate three basic features: a patch of soil, a constructed canal, and a vegetated area. This type of real world data set inspired this line of research since the ability to create aggregated groups from abundant 3D point data, separating distinctly different features in a scene, would be helpful, efficient, and revealing. If the data collected from the three features in this scene were segmented into separate groups, then theoretically the characteristics, variable sets, clustering methods, and parameters may be extended to geographic features of larger extents.

First, the Archaeological Site LiDAR scan was sampled for computational efficiency. Then, the scan was labeled as 3 groups: the grass, the stone canal, and the soil. Figure 29 shows images of the unsampled data, the sampled data in RGB color, and the sampled data colored per their reference groups.

Archaeological Site LiDAR (Unsampled)



Archaeological Site data (Sampled)

Archaeological Site data (Reference groups)



Figure 29. Archaeological Site LiDAR scan, sampled data, and labeled reference groups.

After the Mean Shift clustering was complete, the accuracy assessment produced segmentation performance curves. Four of these are shown in Figure 30 and are examples of higher performing variable sets. Three of the four images show that all three of the reference groups were correctly segmented through the 0.6 tolerance level, meaning that the three output groups had between 60-70% of their respective reference groups' points. The XYZ-MV-RGB variable set maintained two groups with above 70% of their respective reference groups.



Figure 30. Archaeological Site LiDAR segmentation performance curve examples.

The Max. AUC and Avg. AUC rankings, shown in Table 23, show the highest performing variable sets for one bandwidth and across all bandwidths, respectively. Variable sets including RGB had the highest performing bandwidths, probably due to the high separability of values for the green grass. However, the alternative color spaces like AB, UV, and LAB performed well overall. The MV did not perform well, except where it was combined with the XYZ set. The spatial location XYZ variables may not have performed well over all of the tests, but were included in several of the highest performing bandwidths.

Max AUC	Variable Set	Avg AUC	Variable Set
0.50	RGB	0.32	AB
0.50	XYZ-RGB	0.32	UV
0.50	XYZ-MV-RGB	0.32	LAB
0.44	LUV	0.30	RGB
0.44	XYZ-LUV	0.30	LUV
0.44	XYZ-MV-LUV	0.24	XYZ-RGB
0.40	LAB	0.22	XYZ-LAB
0.36	XYZ-LAB	0.22	XYZ-AB
0.36	MV-RGB	0.22	XYZ-LUV
0.34	UV	0.20	XYZ-UV
0.34	AB	0.14	MV-RGB
0.34	XYZ-UV	0.14	XYZ-MV-LAB
0.34	XYZ-AB	0.14	XYZ-MV-RGB
0.34	MV-LUV	0.14	XYZ-MV-LUV
0.34	MV-LAB	0.14	MV-LAB
0.34	MV-UV	0.14	MV-LUV
0.34	MV-AB	0.14	MV-AB
0.34	XYZ-MV-LAB	0.14	MV-UV
0.34	XYZ-MV-UV	0.14	XYZ-MV-AB
0.34	XYZ-MV-AB	0.12	XYZ-MV-UV
0.20	XYZ	0.02	XYZ
0.00	MV	0.00	MV
0.00	XYZ-MV	0.00	XYZ-MV

Table 23. Archaeological Site LiDAR Max. AUC and Avg. AUC values.

The segmentation surface plots reveal more clearly the characteristics of the previous results. The 6 selected images are shown in Figure 31. The top three show that the RGB, the XYZ-RGB, and the XYZ-MV-RGB achieved three groups having over 60% of their respective reference groups. They achieve the highest performance at a few bandwidths. The bottom three images show how persistent some of the variable sets were over all of the bandwidths. The thematic variables showed ability to segment the data as previously in the other case study. The difference here is that the RGB did well, probably due to the highly separable set of data representing green grass. Finally, when the XYZ variables were combined with color variables, the range of effective bandwidths was reduced and the performance for limited bandwidths was increased,



Figure 31. Archaeological Site LiDAR segmentation surface plot examples.

The 3D perspective plots again illustrate which points were grouped together. In Figure 32, there are 6 images showing the effects of using different spatial location, spatial relationship, and thematic variable sets. The results show segmentation characteristics that were similar to previous experiments, though they offered an opportunity to discuss the methods as applied to landscape features.





Figure 32. Archaeological Site LiDAR 3D perspective plot examples.

First, the XYZ-RGB separated different features into small localized patches. With a relatively small bandwidth of 0.2, small groups were created. Since the scene was divided into many groups of nearby points having similar thematic attributes, it could be considered over-segmented. The XYZ-MV-LAB variable set, having a small bandwidth of 0.2, again broke up the scene into small groups, though the MV reduced the number of points included in groups overall. The UV variable set drastically generalized the scene; the majority of points were in two groups. Without the lightness variable, the scene was segmented based upon the two dimensions of color. This was effective at separating the green grass from the brownish soil and rocks. With

the lightness included in the LUV variable set, the landscape was generalized into three large groups of points. The soil and rocks were mostly separated since the rocks were generally darker. However, working with only spectral values, the clustering grouped the lighter tones on top of the rocks together with the sandy soil. The XYZ-LUV and XYZ-RGB variable sets demonstrated how local proximity was helpful in separating the tops of rocks from the soil. The XYZ component did separate the grass into a several groups, though there were fewer due to a larger bandwidth of 0.4. Also, upon close inspection of the results, the separate group at the top of the grassy area was actually brownish in the original LiDAR scan indicating a small sliver of soil. This nuance in the data was identified with the aid of the cartographic generalization process implemented with the Mean Shift clustering. Ultimately, both the XYZ-LUV and the XYZ-RGB variable sets generalized the scene enough to separate different features in the landscape. Not too many groups were created and some could be combined for subsequent processing. The XYZ-RGB variable set appeared to have a slightly more refined output as it completely separated the rocks from the soil. Features were grouped per their spatial and thematic characteristics and the complex landscape was simplified.

Chapter 5 - CONCLUSIONS

5.1 Overview

To conclude this thesis, I address several previously stated propositions. The main topics are organized in the following sections: two general hypotheses, five sets of experiments, contributions, and future research.

5.2 Two General Hypotheses

High level conclusions drawn from my work are related to my two general hypotheses. First, this research supports the idea that groups of data representing geographic phenomena in the environment can be identified using spatial location, spatial relationship, and thematic characteristics with the Mean Shift clustering technique. Second, combinations of variable sets enhance the ability to find such phenomena compared to single variable sets.

5.3 Five Sets of Experiments

The five sets of experiments listed in the Methods chapter each provided insights to the process of aggregating thematically attributed 3D data. The knowledge gained from developing controlled experiments with synthetic data guided the successful application of the methods to the real-world terrestrial LiDAR data. Notable results are listed in the following five paragraphs.

The first evaluation compared the K-means and Mean Shift clustering methods. It illustrated that benefits to using the Mean Shift clustering are that it is more flexible regarding the number of output groups and had a higher overall performance. This offered initial evidence for the potential success of clustering multivariate geographic data with the Mean Shift technique. The second comparison, regarding spatial resolution, showed that a potentially under-sampled low-resolution data set greatly degraded the grouping of points. However, the high-resolution data set, with considerably more points, did not improve performance compared to the baseline data set. This suggests that only a moderate amount of data (in relation to the bandwidth of evaluation) is required to find groups of points representing features in the scene.

The third evaluation studying the effects of noise provided evidence that even a reasonably small amount of noise in the data can significantly degrade the segmentation results. A benefit of the noise evaluation was that it allowed for the generation of multiple synthetic realizations of the data which provided a way to evaluate the performance of different variables through statistical testing. The results showed that combinations of variable sets were found to perform higher than individual variable sets. In comparing the color variable sets in the synthetic data, the RGB had higher performance. However, this was ultimately found to be influenced by the structure of the synthetic data, which had higher separability for the RGB set as compared to the LUV and the LAB sets. The MV spatial relationship variable set was found to perform poorly with regard to the measures of the accuracy assessment, though this variable may provide other benefits for separating planar and non-planar regions. The spatial location (XYZ) and thematic (RGB) variable sets both showed high performance when included in the synthetic data evaluation.

The fourth evaluation involved an alternative color configuration in the synthetic data. It resulted in spatially disjoint sets of points with similar colors being grouped together This indicated that the thematic variable set had a strong influence on the clustering results. Again, the

structure of the synthetic thematic data was more separable than either the spatial location or spatial relationship.

Finally, the real world case studies of the Indoor and Archaeological Site LiDAR data provided the most encouraging results. They demonstrated that the variable sets, clustering techniques, and parameters can have success in grouping points representing simple objects and landscape features. Supplementing the spatial data with the appropriate thematic information allowed for features to be grouped while alleviating the effects of shadows and environmental lighting. Including spatial and thematic data improved the cartographic generalization process and aggregating real-world point data effectively simplified the scene. This ultimately resulted in a higher-level of ontological information about features in the real-world scenes.

5.4 Contributions

Relating to my contributions listed in the introduction, I've reached a variety of conclusions. Though the unstructured nature of the data, without topological connectivity or relationships, presented difficulty, the <u>3D vector point data</u> structure was a good match for the Mean Shift clustering technique. This is especially so considering multiple thematic attributes can be associated with the spatial locations.. The <u>Mean Vector</u> variable, a variable derived from local points in a spatial neighborhood provided insufficient information to fully represent neighborhood relationships. This variable only managed to help group all points representing planar regions together, with edges, corners and areas of high variability each grouped separately. It is noteworthy to mention that, although the Mean Vector degraded results for this accuracy assessment, it also shows potential for other applications. The stress was concerned with finding points representing homogeneous regions. If, on the other hand, an aim is to find

edges, corners, or areas of high variability in contrast to planar regions, then the Mean Vector show potential. It could be used in edge-based segmentation processes, possibly by finding sets of points that could be used to generate vectors representing *the edges* of homogeneous regions.

The evaluation of *color spaces* provided two main insights. First, this common type of thematic attribute greatly aided the grouping of points and helped separate features in close proximity with similar Mean Vector values. Second, the LUV and LAB variable sets both allowed lightness to be removed and showed success in alleviating the degrading effects of shadows and environmental lighting in the process of segmenting real-world data. However, an important note is that when the segmentation used only the UV and AB color variables for handling shadows, white and black features were indistinguishable as they vary primarily in lightness. Generally, the LUV and LAB color spaces can help to separate colors into separate clusters. However, the RGB variable set effectively segmented the Archaeological Site LiDAR scan since the grass was particularly greener than the rest of the scene.

The <u>Mean Shift</u> clustering technique provided a flexible and generally successful method for clustering similar points in a multi-variate data set. This technique, used often in the discipline of Computer Vision, showed potential for segmenting real-world geographic data having three spatial dimensions and thematic attributes. A general guideline regarding the Mean Shift process is that the bandwidth distance of evaluation should be carefully selected to find different scales of features. This will guide the process in finding different sizes of clusters in variable space. If smaller bandwidths are used, smaller clusters may be found as the bandwidth will not bridge a

large gap between clusters. If bigger bandwidths are used, the results will eventually show one large, under-segmented, cluster.

The <u>segmentation performance accuracy assessment</u> included elements from traditional classification assessments and image analysis disciplines. These elements had previously been applied to range image segmentation research and, as such, were suitable for the terrestrial LiDAR data in this investigation. They could be very beneficial for geographic data analyses in general.

My extensions of the previous accuracy assessment framework were key to fully evaluating the numerous variables, methods, and parameters. The *area under the curve indices* condensed the information from the segmentation performance curves. The Max. AUC values provided a single metric for determining the highest performing Mean Shift bandwidths. A second metric, the Avg. AUC, provided a measure of a variable set's persistence in performance throughout all bandwidths. The Avg. AUC also provided an overall performance measure for the statistical analyses, helping to identify the most effective variable set combinations. The *segmentation performance surface plots* were complementary to the AUC metrics and particularly helpful for understanding the various levels of performance for each variable set through all of the Mean Shift bandwidths. Finally, the 3D perspective plots, common for visual evaluation, revealed the which points were grouped in relation to the features they represent.

5.5 Contributions to Theory

As a contribution to theory, I combined elements of two geographical concepts, *cartographic generalization and geographic ontology*, formalizing a particular process and method of

geographic representation. Two specific theories were discussed, the cartographic generalization of McMaster and Shea (1992), and an ontological hierarchy from Couclelis (2009). These two theories, though they came from separate threads of geographic inquiry, are very parallel in relation to this investigation. Both are related to geographic representation of features in the landscape and both present processes that similarly combine information to better understand those features. The generalization process of aggregation was a transformation that grouped 3D points into representations of areal regions having similar spatial and thematic characteristics (McMaster & Shea, 1992). The ontological hierarchy also developed simple objects from spatially connected, homogeneous points with similar (thematic) observables (Couclelis, 2009). These two theories ease interpretation of complex data sets by similarly developing groups of points that represent recognizable features in the landscape.

To identify features in the landscape, a well-defined ontology is necessary. Certain properties and relationships must be understood, clarified, and explicitly modeled with the data. An example from this thesis that could be improved is the interaction between the Mean Vector and the accuracy assessment. While the Mean Vector may have fulfilled the purpose of separating planar regions from non-planar regions and is potentially helpful for other applications, it reduced the number of points per region and reduced the performance with respect to the accuracy assessment.

Another example of refining an ontological model in this thesis relates to the relationship between the distribution of spatial or thematic data values, the level of precision, and the Mean Shift bandwidth of evaluations. The bandwidth distance should be selected to guide the

clustering process in finding certain sizes of clusters. For instance, small features may be found with small bandwidths. This could apply to either the spatial or thematic data. This also relates to the level of precision of data collection. The ability to find features is dependent upon the number of data points collected. If more data points were used to represent the blue cup, yellow ball, or yellow ducks in the Indoor data set, they would have composed a sufficiently dense cluster in variable space to be grouped by the Mean Shift. The red toolbox and blue pail show that increasing spatial resolution relative to the size of the feature can improve clustering performance. Collecting a sufficient amount of data to identify a feature could be an important component of the domain ontology. The selection of variables that relate to particular characteristics can be guided by their distribution of values. For instance, the Archaeological Site scan benefitted from the distribution of green values, resulting in segmentation of the grassy areas. It didn't have as many colors or shadows as the Indoor scan, which had better segmentation with the alternative color spaces. A specific definition of ontology of characteristics and relationships for segmenting desired features is evidently important in cartographically generalizing thematically attributed 3D data.

5.6 Future Research

Considering future directions from this investigation, three main topics that come to mind. First, this thesis extended an evaluation of experimental methods to real world LiDAR data. Applying the methods to a larger extent would allow for the selection of points representing larger geographic features. A mosaic of terrestrial LiDAR scans or a data generated through fusion of aerial LiDAR and photographs could be used. A second topic would be to incorporate spectral reflectance with more variation or in different bandwidths, such as the near infrared or hyperspectral data. The third main topic for future investigation would be to conceptualize and

employ a better 3D neighborhood relationship metric than the Mean Vector; it only separated planar regions from other configurations. A more complete metric would identify different planar orientations or possibly address regions of similar curvature. Overall, this thesis demonstrated how diverse techniques and theories from other disciplines can be integrated with those of Geography and in the process developed new approaches for indentifying geographic features represented by richly attributed 3D data.

Applications that this research could potentially benefit are varied. Spatial and thematic data representing geographic features in natural or urban settings could be segmented. The methods may help to select and quantify vegetation. Models of under-story vegetation could complement canopy information. Landforms such as dunes, outcrops, and caves could be segmented from high resolution data. Data relating to different constituents of soils, water bodies, or the atmospheric could be mapped in 3D. The potential for studying objects at different scales, such as atomic, biophysical, or astronomic, and the movement of such objects could benefit from 3D feature segmentation. Finally, this study can provide information for basic research in 3D spatial analysis and uncertainty which aid cartographic representation of regions and their boundaries.

REFERENCES CITED

REFERENCES CITED

- Agarwal, P. (2005). Ontological considerations in GIScience. *International Journal of Geographic Information Science*, 19 (5), 501-536.
- Ball, G. H., & Hall, D. J. (1967). A clustering technique for summarizing multivariate data. *Systems Research and Behavioral Science*, *12* (2), 153-155.
- Barnea, S., & Filin, S. (2008). Segmentation Of Terrestrial Laser Scanning Data By Integrating Range And Image Content. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVII, Part B5*, pp. 745-750. Beijing.
- Comaniciu, D., & Meer, P. (2002). Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34 (5), 603-619.
- Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of the Environment*, 37, 34-46.
- Couclelis, H. (2009). Ontology, Epistemology, Teleology: Triangulating Geographic Information Science. In *Research Trends in Geographic Information Science* (pp. 3-15). Berlin: Springer Verlag Lecture Notes in Geoinformation and Cartography.
- Couclelis, H. (1992). People Manipulate Objects (but Cultivate Fields)-Beyond the Raster-Vector Debate in GIS. In A. U. Frank, I. Campari, & U. Formentini (Eds.), *Theories and Methods in Spatio-temporal Reasoning in Geographic Space* (Vol. 639, pp. 65-77). New York: Springer Verlag Lecture Notes in Computer Science.
- Foody, G. M. (2002). Status of Land Cover Classification Accuracy Assessment. *Remote* Sensing of Environment, 80, 185-201.
- Frank, A. U. (2000). Geographic Information Science: New methods and technology. *Journal of Geographical Systems*, 2, 99-105.
- Georgescu, B., Shimshoni, I., & Meer, P. (2003). Mean shift based clustering in high dimensions: A texture classification example. *Ninth IEEE International Conference on Computer Vision* (pp. 456-463). Nice, France: IEEE Computer Society.
- Goodchild. (1990). Geographical Information Science. *Keynote Address*. Zurich: Fourth International Symposium on Spatial Data Handling.
- Goodchild, M. F., & Wang, M. -H. (1989). Modeling error for remotely sensed data input to GIS. *Proceedings, AutoCarto 9* (pp. 530-537). Falls Church, VA: ASPRS/ACSM.

- Goodchild, M. F., Yuan, M., & Cova, T. J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science*, 21 (3), 239-260.
- Guarino, N. (1998). Formal Ontology and Information Systems. *Formal Ontology and Information Systems (FOIS)* '98 (pp. 3-15). Trento, Italy: IOS Press.
- Haralick, R. M., & Shapiro, L. G. (1985). Image Segmentation Techniques. *Computer Vision, Graphics, and Image Processing*, 29 (1), 100-132.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28 (1), 100-108.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer.
- Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P. J., Bunke, H., Goldgof, D. B., et al. (1996).
 An Experimental Comparison of Range Image S eg m e n t at i o n AI go r i t h m s. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18 (7), 673-689.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. ACM Computing Surveys, 31 (3), 264-323.
- Keates, J. S. (1973). Cartographic Design and Production. New York: John Wiley & Sons, Inc.
- Ledoux, H., & Gold, C. M. (2008). Modelling Three-Dimensional Geoscientific Fields with the Voronoi Diagram and its Dual. *International Journal of Geographical Information Science*, 22 (5), 547-574.
- Lee, I.-C., Wu, B., & Li, R. (2009). Shoreline Extraction from the Integration of Lidar Point Cloud Data and Aerial Orthophotos using Mean Shift Segmentation. *American Society of Photogrammetry and Remote Sensing Annual Conference*. Baltimore, Maryland.
- MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). Berkeley: University of California Press.
- McMaster, R. B., & Shea, K. S. (1992). *Generalization in Digital Cartography*. Washington D.C.: Association of American Geographers.
- Ortiz, A., & Oliver, G. (2006). On the Use of the Overlapping Area Matrix for Image Segmentation Evaluation: A Survey and New Performance Measures. *Pattern Recognition Letters*, 27, 1916-1926.
- Paris, S., & Durand, F. (2007). A Topological Approach to Hierarchical Segmentation using Mean Shift. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota: IEEE.

- Pequet, D. (1988). Representations of Geographic Space: Toward a Conceptual Synthesis. Annals of the Association of American Geographers, 78 (3), 375-394.
- R Development Core Team. (2009). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Schwering, A. (2008). Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. *Transactions in GIS*, *12* (1), 5-29.
- Shimshoni, I., Georgescu, B., & Meer, P. (2006). Adaptive Mean Shift Based Clustering in High Dimensions. In G. Shakhnarovich, T. Darrell, & P. Indyk (Eds.), *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice* (pp. 203-220). Cambridge, Massachusetts: MIT Press.
- Tou, J. T., & Gonzalez, R. C. (1974). *Pattern Recognition Principles*. Reading, Massachusetts: Addison-Wesley.
- Weibel, R. (1997). Generalizations of Spatial Data: Principles and Selected Algorithms. In M. van Kreveld, J. Nievergelt, T. Roos, & P. Widmayer (Eds.), *Algorithmic Foundations of Geographic Information Systems* (Vol. 1340, pp. 99-152). New York: Springer Verlag Lecture Notes in Computer Science.
- Xu, R., & Donald Wunsch, I. (2009). *Clustering*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Yao, W., Hinz, S., & Stilla, U. (2009). Object Extraction Based on 3D-Segmentation of Lidar Data by Combining Mean Shift with Normalized Cuts: Two Examples from Urban Areas. Urban Remote Sensing Joint Event. Shanghai, China: IEEE Geoscience and Remote Sensing Society.
- Zhang, Y. J. (1996). A Survey on Evaluation Methods for Image Segmentation. *Pattern Recognition*, 29 (8), 1335-1346.
- Zheng, L., Zhang, J., & Wang, Q. (2009). Mean-Shift-Based Color Segmentation of Images. Computers and Electronics in Agriculture, 65, 93-98.