2009

Feature extraction from high resolution digital elevation models of Mars

Leslie E. Keely-Meindorfer
San Jose State University

FEATURE EXTRACTION FROM HIGH RESOLUTION DIGITAL ELEVATION

MODELS OF MARS

A Thesis

Presented to

The Faculty of the Department of Geography

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Leslie E. Keely-Meindorfer

August 2009

UMI Number: 1478590

UMI®

Dissertation Publishing

ProQuest®
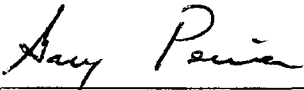
SAN JOSE STATE UNIVERSITY

The Undersigned Thesis Committee Approves the Thesis Titled

Feature Extraction from High Resolution Digital Elevation Models of Mars

by
Leslie E. Keely-Meindorfer

APPROVED FOR THE DEPARTMENT OF GEOGRAPHY

_____   15-June-2009
Dr. Richard Taketa,   Department of Geography   Date

_____   6/1/09
Dr. Gary Pereira,   Department of Geography   Date

_____   5/26/2009
Dr. Paula Messina,   Department of Geology   Date

APPROVED FOR THE UNIVERSITY

_____   7/20/09
Associate Dean,   Office of Graduate Studies and Research   Date

ABSTRACT

FEATURE EXTRACTION FROM HIGH RESOLUTION DIGITAL ELEVATION MODELS OF MARS

by Leslie E. Keely-Meindorfer

This work examines the process of extracting landscape features from high resolution digital elevation models. These models were created with images from a stereo camera resident on a satellite orbiting the planet Mars. Previous work in this context describes the use of classification and image segmentation techniques for feature extraction. Several of these techniques are examined and discussed and an approach is presented. This approach employs cluster analysis, image segmentation, and scale-space to create areas of morphological homogeneity. Subsequently, these areas are characterized and merged to define features commonly found in the Martian landscape.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

One of the primary tasks in a mission to the planet Mars is to determine the landing site. The process of selecting the landing sites for the Mars Exploration Rovers (MER) mission lasted for more than two years and involved a broad cross-section of scientists and engineers from the planetary science community. They looked at the characteristics of approximately 155 candidate sites before settling on one for each rover. The primary objective was to meet all engineering requirements in order to ensure a safe landing. A secondary objective was that of obtaining the highest science return. Overall, 19 engineering and 14 science criteria were examined. Many of these criteria were spatial and morphological in nature. For example, surface gradient, rock abundance, elevation, trafficability, and winds were important. Undoubtedly, this task would have been facilitated by the availability of high resolution topographic and geologic maps.

Mars Reconnaissance Orbiter (MRO) is a relatively new satellite capable of providing the data necessary to create maps that cover the entire planet at 1 meter scale, the highest resolution Mars orbital data to date. This onslaught of information presents difficulties for traditional methods of mapping and feature identification and can be more easily stored, managed, and analyzed with the aid of a geographic information system (GIS). This GIS would contain information about each landform discovered on the surface of the planet. These landforms and their attributes would be available for search,

classification, and analysis. Spatial attributes (e.g., location, dimension, and adjacency), morphological attributes (e.g., landform type), spectral attributes (e.g., color), and results of scientific models as well as annotation would be stored with these objects.

An initial step toward the creation of such a GIS is the transformation of orbital data into *landform elements*. Each of these elements consists of an area of homogeneity that represents a morphological feature on the surface of Mars and provides the basis for building the GIS. The objective of this work is to examine the process of using feature extraction from high resolution digital elevation models (DEMs) to create landform elements. The DEMs are derived from images taken by a stereoscopic camera carried on MRO. The extraction method is based on classification, image segmentation, and scale-space. Once the elements are extracted from the data, they are used to construct simple landform objects that could be stored in a GIS.

CHAPTER 2: BACKGROUND

Mars Reconnaissance Orbiter

MRO is a satellite sent by the National Aeronautics and Space Administration (NASA) to Mars to study the history of water on the planet. It contributes to the NASA science goals of determining whether life ever arose on Mars, characterizing the climate and geology of Mars, and preparing for human exploration. The spacecraft achieved orbit around the planet in 2006 and has since sent over 2 terabytes of data back to Earth. MRO carries eleven instruments among which are three cameras, a spectrometer, a radiometer, and radar. These science instruments provide data for the purposes of global mapping, regional surveying, and high-resolution targeting. Two of the cameras, High Resolution Imaging Science Experiment (HiRISE) and Context Camera (CTX) are used to produce DEMs. HiRISE operates at a high enough resolution to distinguish 1 meter size objects and produces stereo pairs with its stereo lens option. These stereo pairs are then used to create the DEMs. CTX is aligned with HiRISE and provides a broader contextual view (a pixel is approximately 10 m x 10 m). Although it does not have the more accurate stereo camera, DEMs are produced with CTX from multiple passes of the same area. When viewing the DEM, one of the images is typically overlaid on top as a Digital Raster Graphic (DRG) to show color and details. DEMs from the HiRISE

cameras were used in this work because of their high resolution and, as the result of a

stereo camera, should sustain less error.

Digital Elevation Models

A DEM is a regular grid of elevation values called a *raster*. It is similar to an

image except that each pixel is called a *post* and defines an elevation sampled at a point

on the surface. This project uses DEMs created with a stereo camera on MRO. A stereo

camera operates much like stereo vision. In Figure 1 the position of a point *A* on the

ground can be deduced from its position in each image of the stereo pair and the camera

angles and distances. *C1* and *C2* are the left and right cameras and *a'* and *a"* are left

and right image points (pixels) for *A*.
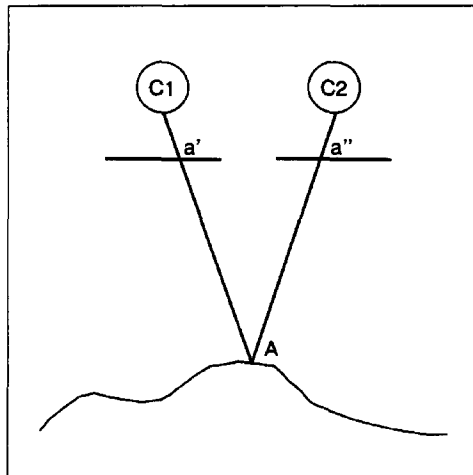
Figure 1. Construction of DEM from stereo image pair.

A DEM is decomposable into several derivatives, which together define the morphological nature of a landform. Relative elevation (e.g., peak or depression), gradient (e.g., gorge or delta), and curvature (e.g., divergent or convergent slope) are computed directly from the DEM. Furthermore, a number of related attributes can be derived from elevation data. Drainage patterns from a hydrological model, a topographic wetness index (TWI) that represents water flow and stagnation across a given area, incident solar radiation, and dominance over the surrounding landscape are all useful quantities that can be derived from an elevation grid. These quantities may rely on other data, such as solar angle, in addition to elevation. DEM derivatives, also known as *attributes*, are used as the basis for defining the homogeneous area that is a landform element.

The high resolution of the MRO images can result in rapid change across the DEM which is often referred to as a *high level of detail* or *high detail*. The term *high relief* indicates a large change in elevation or Z axis of the landscape. In contrast, high detail indicates change in the plane or X and Y axes of the landscape. High detail sometimes causes a problem in feature extraction called *over-segmentation* where the result consists of many very small features. These small objects are generally not usable and require further processing.

Several methods for the creation of landform elements using elevation-derived attributes are described in the literature. Commonly, these techniques developed regions of homogeneity based on the attributes and then classified those regions or groups of regions as elements. These techniques have been applied to study sites on Earth. Some of them have been used to examine low resolution data sets from previous Mars orbiters such as Mars Global Surveyor (MGS) as well. In each technique, a *feature vector* of DEM attributes was created for each post of the DEM. The attributes consisted of elevation and various elevation derivatives. Next, using the feature vector, the posts were combined into homogeneous landform elements. For this work, a number of these methods were examined and some preliminary testing was performed to determine an approach appropriate for high detail, extra-terrestrial data. They can be divided into two groups: segmentation-classification and classification-segmentation.

The first group, those that performed segmentation first, segmented the posts into elements based on homogeneity. Then each segment was classified based on its mean feature vector. The techniques in this group include watershed segmentation, object-based image analysis, and support vector machine.

Watershed Segmentation - Surface feature extraction is of interest in a number of fields. In addition to the spatial needs of geography and geology, engineering and computer graphics require partitioning methods for mesh simplification. Mangan and Whitaker (1999) described a method for partitioning a 3D polygonal surface mesh into

segments for computer graphics and visualization. They used morphological watersheds to segment the total surface curvature of the mesh into patches. This technique was applied to computer generated models with good results. However, these models were relatively smooth and the result degraded significantly when noise was introduced. In tests, this method produced a very over-segmented result when applied to highly detailed Mars terrain data sets and as such, took an inordinate amount of time to process. Consequently, the result did not reflect the nature of the landscape and required substantial additional processing to produce landform elements.

Object-based Image Analysis - This technique provided parameters for controlling over-segmentation. Dragut and Blaschke (2006) described a process for automated landform element classification using object-based image analysis: creating landform elements through image segmentation using a formula for spatial and spectral homogeneity. In their case, the feature vector provided the spectral portion of the formula. Constraints were applied through weighting the two types of homogeneity and by setting an element scale threshold. The spatial homogeneity of the element was further controlled by two sub-weights of smoothness and compactness. The general scale of the segments was determined by the scale threshold, providing a means for increasing or decreasing element size and thus handling the issue of high detail.

This method was applied to two study sites, one a hilly region in the Transylvanian Plain of central Romania and the other in the Berchtesgaden National Park of the German Eastern Alps. The authors used a combination of elevation and derived

attributes of gradient and curvature. The resulting landform elements were classified according to dominance over neighboring elements, gradient, and relative elevation (upland, midland, or lowland). These classification results were directly integrated into a GIS and were found to be reproducible and comparable between the data sets.

In tests on Martian data sets, landform elements produced by this method were homogeneous but often did not reflect the morphology of the landscape. Due to the high detail of the data, changes to the homogeneity and scale parameters were necessary. However, in most cases, these changes produced insignificant or unpredictable results and could not be quantified in terms of the data set.

Support Vector Machine - Stepinski, Ghosh, and Vilalta (2006, 2007) applied machine learning to the task of geomorphic mapping of the Martian surface. They tested two algorithms to segment a 500 m/post DEM based on its elevation-derived attributes: a homogeneity measure and K-means. These methods were applied to 6 different sites on Mars with the aim of identifying crater floors, convex crater walls, convex ridges, concave ridges, and inter-crater plateaus. A training set of manually labeled topographic objects was created and the landform elements resulting from the segmentation step were classified using various techniques including Naive Bayes, Neural Network, Bagging, and Support Vector Machines (SVM): the SVM being the most successful. This method produced good results with accuracy estimated between 78 and 85 percent. They found that the success of their technique depended on the quality of segmentation and the choice of an appropriate feature vector.

The second group, those that classify first, classified each post according to its feature vector. Then, neighboring posts of the same class were combined to form a landform element. These methods include ISODATA Unsupervised Classification, Segmentation Using Heuristic Rules and Fuzzy Logic, Fuzzy K-means Classification, and Self Organizing Map.

ISODATA Unsupervised Classification - Ventura and Irwin (2000) created landform elements using the ISODATA method, a modified K-means technique. The modification allows for the automatic determination of the final number of clusters through splitting and merging. Once the posts were classified to clusters, elements were created by aggregating similar neighboring posts. The input to ISODATA required some processing since it must be at least roughly Gaussian. Additionally, input parameters for the maximum number of clusters, and the thresholds where splits and merges should occur were required. The thresholds were determined by the standard deviation of the data.

Ventura and Irvin (2000) examined this technique for defining landform elements in a classification process for soil studies and compared it to manual methods. They tested ISODATA using the attributes gradient, aspect, curvature, TWI, and incident solar radiation computed from DEMs acquired for a study site in Pleasant Valley, Wisconsin. They found that the results of the automated method could aid those interested in soil-landscape studies but the feature vector should be chosen with the nature of the landscape and the phenomena being studied in mind. Additionally, they found that where manual

classification methods are subjective, numerical methods, while requiring some subjective input, are largely objective and repeatable.

In tests on Mars data, the data were converted to Gaussian by applying a log10 function to those attributes that warranted it. The parameters required for controlling splits and merges greatly affected the resulting number of clusters. Although more automatic than others, this method required substantial preprocessing and knowledge of the statistical nature of the data sets.

Segmentation Using Heuristic Rules and Fuzzy Logic - MacMillan, Pettapiece, Nolan, and Goddard (2000) described a procedure for determining landform elements for use in precision farming in Alberta, Canada. The problem this work addressed is one of the production of a generic classification of landform elements that can be applied to a wide variety of landscapes in an automated fashion without alteration. The method employed elevation-derived attributes, a hydrological model, heuristic rules, and fuzzy logic. The hydrological model provided the location of each post relative to the nearest stream or divide. Constraints were provided in the definition of the heuristics and bounds of the fuzzy classes. A complex classification scheme was developed, a pixel level classification was performed, and the results aggregated into 15 landform element types. The inclusion of several measures of relative and absolute landform position reduced spatial fragmentation. This technique resulted in landform segmentation model procedures that were considered generic enough to be applied to a wide variety of agricultural landscapes.

This method relied heavily on hydrological information and required an existing and somewhat complex classification. It was very comprehensive but depended substantially on models and expert knowledge that are available for Earth but not yet for Mars.

Fuzzy K-means Classification - In three works Arrell, Fisher, Tate, and Bastin (2007), Burrough, van Gaans, and MacMillan (2000), and Ventura and Irwin (2000) used a K-means method extended with fuzzy logic to define landform elements. K-means is essentially a sorting and binning procedure where an expert determines the number of bins and the rules for sorting. The standard K-means procedure sorts a set of data points into a specified number of bins or *clusters* based on the feature vector for each data point. A fuzzy version of K-means was used by these authors with DEM posts serving as data points. Instead of assigning each post to a single cluster, fuzzy K-means assigned each post a membership (0 - 1) in each cluster; the memberships for a single post summing to 1. After the cluster analysis was complete, adjacent posts of the same cluster were combined to form landform elements. Burrough et al. calculated a *Confusion Index* (CI) for each post. This is the ratio of the first sub-dominant membership value to the dominant membership value. Posts with a CI that exceeds a certain threshold did not fall predominantly into any single cluster and were labeled as *intragrades*. Intragrades were not considered to be part of a landform element allowing some posts to remain unclassified. As with K-means, fuzzy K-means is an iterative process and continued until a convergence threshold was reached.

Burrough et al. (2000) performed landform classification using fuzzy K-means on a high-resolution DEM data set covering an area of farmland in Alberta, Canada, and one covering a region of the French pre-Alps. The two sites contrasted in that one was a moderately rolling landscape whereas the other was one of strong relief. Their work used the attributes gradient, profile curvature, plan curvature, total annual incident solar radiation, TWI, and distance from ridge lines. Additionally, it provided methods for overcoming artifacts introduced in the derivative computation and for handling large data sets through sub-sampling. The results indicated success in producing a useful and meaningful division of the landscape.

Arrell et al. (2007) examined the scale dependency of morphometric classes of a study site in Snowdonia, Wales using a DEM of the region. A fuzzy K-means classifier was applied to elevation-derived attributes at several resolutions and class persistence was evaluated at each resolution. Resolution levels were obtained by sub-sampling the primary DEM. They found that a 50 m resolution data set was too sensitive to surface noise and impeded the success of morphometric identification. They also found that fuzzy K-means cluster analysis provided a classification with high information content. Finally, they concluded that the resolution of the data set would determine the morphometric classes that would be identified and that extreme morphometric classes such as ridges were found to persist in the same geographical space throughout all resolutions.

Choice of feature vector was based on expert knowledge of the landscape. Constraints included the number of clusters, the convergence threshold, the fuzziness factor (typically 1.5), and the intragrade threshold. Test results on Mars data indicated that the most effective values for these parameters were consistent with the literature. The resulting elements reflected the morphology of the terrain and no data-specific parameters were necessary. Although it was more time-consuming, the requirements for fuzzy K-means tended to be simple and unbiased.

Self Organizing Map - Bue and Stepinski (2006) developed a method for automated classification of Martian landforms using a self-organizing map (SOM) technique. They classified DEM posts based on a feature vector of elevation, gradient, flood, flooded gradient, contributing area, and flooded contributing area from low resolution MOLA DEMs. The classified posts were then combined into elements using Ward's agglomerative hierarchical clustering method. Ward's method is an iterative cluster comparison and merge based on minimum information loss.

The data for their work was a DEM from the Mars Orbiter Laser Altimeter (MOLA) onboard the Mars Global Surveyor spacecraft. A laser altimeter is an instrument that employs a laser to measure the distance from the spacecraft to the ground. The MOLA resolution was 128 posts/degree. Elevation, gradient, flood, and contributing area (drainage area) were used in the SOM for study sites in the Terra Cimmeria, Reull Vallis, and Margaritifer Sinus regions. The flood field consisted of zero values at each post except for those located inside topographic basins. The contributing

area of a given elevation was the area of posts located above it in a drainage hierarchy. The thematic map resulting from this work consisted of 19 classes of highlands, craters, lowlands, high relief, and channels. Additionally, a "landform abundance" vector was created to provide a quantitative account of the topographic relatedness between different sites.

The work described in the literature was performed on data sets typically with lower resolutions than that of the MRO DEMs and over-segmentation was not a significant issue. In tests on MRO data, the methods in the segmentation-classification group had the advantage of controlling detail by creating elements first. However, typically the elements did not reflect the nature of the landscape and required combination into more meaningful elements. The methods in the classification-segmentation group suffered more from high detail issues such as over-segmentation but resulted in more meaningful elements.

The high detail and small scale of the MRO data could present a problem in detecting the desired landform elements. Changing the scale of the data through sub-sampling and/or filtering is an option for reducing detail. However, this approach will cause arbitrary loss of data. Another approach is to use a *scale-space*. Lindeberg (1996) described scale-space theory as a framework for handling image objects at multiple scales. Scale-space was developed by the computer vision community in response to the fact that real-world objects exist in meaningful form over a specific range of scale. Similarly, Arrell et al. (2007) found that extreme morphometric classes persisted

throughout a range of DEM resolutions. A scale-space is a representation of a data set at multiple scales created by applying a Gaussian blur at a selected kernel radius t, referred to as the *scale parameter*. At each level, objects of a size less than $\sqrt{t}$ are removed while preserving the strongest features and minimizing loss of data.

Two basic trends emerge from the methods described above, *predetermined* and *self-selecting*. Methods of the first combined the feature vector, results of hydrological and other models, and substantial data-specific knowledge, to classify each post and then aggregated the posts into landform elements. Methods of the second trend allowed the landscape to determine the classification, requiring little expert knowledge except in the choice of the attributes for the feature vector. Thus the landform elements essentially "appeared" from the landscape.

In summary, previous work indicates that the combination of image segmentation and classification can be used successfully to define landform elements from DEMs. Additionally, the application of scale-space can be used to control the level of detail in high resolution data. Together, these techniques can be applied to the DEMs of the MRO satellite to create Martian landform elements.

Study Sites

MRO HiRISE DEM data sets are publicly available from the United States Geological Survey (USGS) for the Mars Science Laboratory (MSL) mission candidate

landing sites as well as sites from the Phoenix Mars Lander and MER missions. Each post in these data sets cover an area of approximately 1 m x 1 m and elevation units are meters. One MER site and two MSL candidate sites were selected for this study. The MER site is located in the northern half of the Columbia Hills (Figure 3), named to memorialize the fallen shuttle, Columbia. This site is located approximately 2.5 km southeast of the Spirit landing site in Gusev Crater, which is possibly an ancient lake bed. The hills are composed of rock outcrops much older than the plain of Gusev Crater and are considered to have undergone significant alteration by water. Figure 2 shows a labelled view of the hills from Spirit's landing site. The second site (Figure 4) is located in the eastern portion of the proposed landing ellipse for MSL in Gale Crater. Gale Crater has a complex history of erosion and deposition and houses a very large layered mound at its center. The study site is at the base of the mound in a dune field. The last site (Figure 5) is located in the second landing ellipse in the Mawrth Vallis MSL candidate location and is believed to contain clay-bearing light colored layered bedrock. This site has a rich mineral diversity which could signify a variety of processes of formation. Layering and polygonal fractures are also visible in images of this region.

Compared to the rest of the planet's terrain, these sites appear to be somewhat benign. Safety is of primary importance in landing site selection and a gentle landscape is typically chosen to ensure the successful arrival of the spacecraft. Each site has its own characteristics. Columbia Hills presents hills and a substantial number of craters. Gale Crater features dunes and relatively smooth terrain. Mawrth Vallis, by contrast, has

a few sharp peaks and steeper slopes. Table 1 summarizes the characteristics of the three

study sites. The mean gradient indicates mostly gentle slopes. However, all the sites

have some steep topography. This characteristic is found in the lower slopes of the

Columbia Hills, the slipfaces of the Gale Crater dunes, and the peaks of Mawrth Vallis.
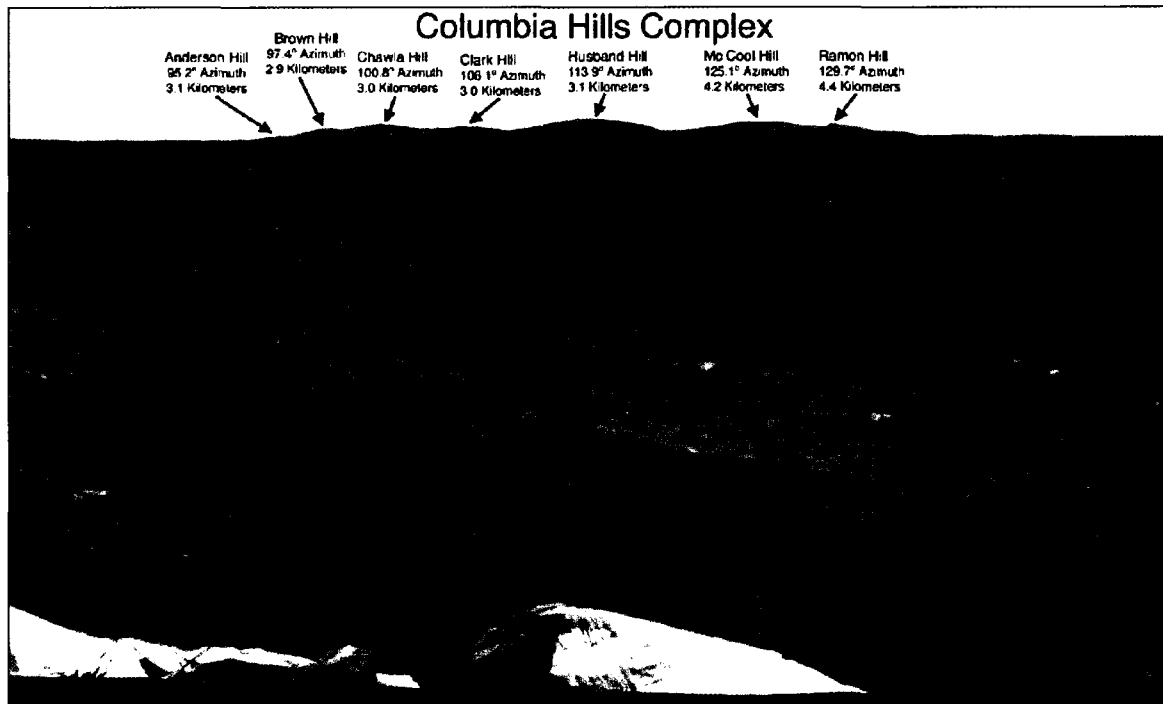
Figure 2. View of Columbia Hills from Spirit's landing site (courtesy JPL/NASA).

Table 1

Study Site Characteristics

Name	Location	Elevation Change	Mean Gradient	Mean Elevation	Terrain
Columbia Hills	14.58 S, 175.52 E	94 m	6.38	-1947.6	Hills and Craters
Gale Crater	4.49 S, 137.42 E	83 m	5.12	-4498.3	Dunes
Mawrth Vallis	24.01 N, 341.03 E	124 m	6.38	-2450.1	Rough, sharp peaks
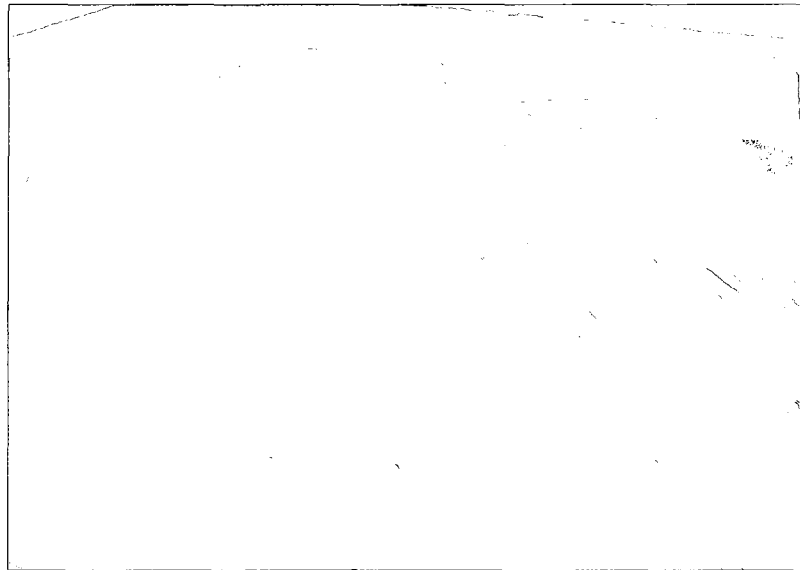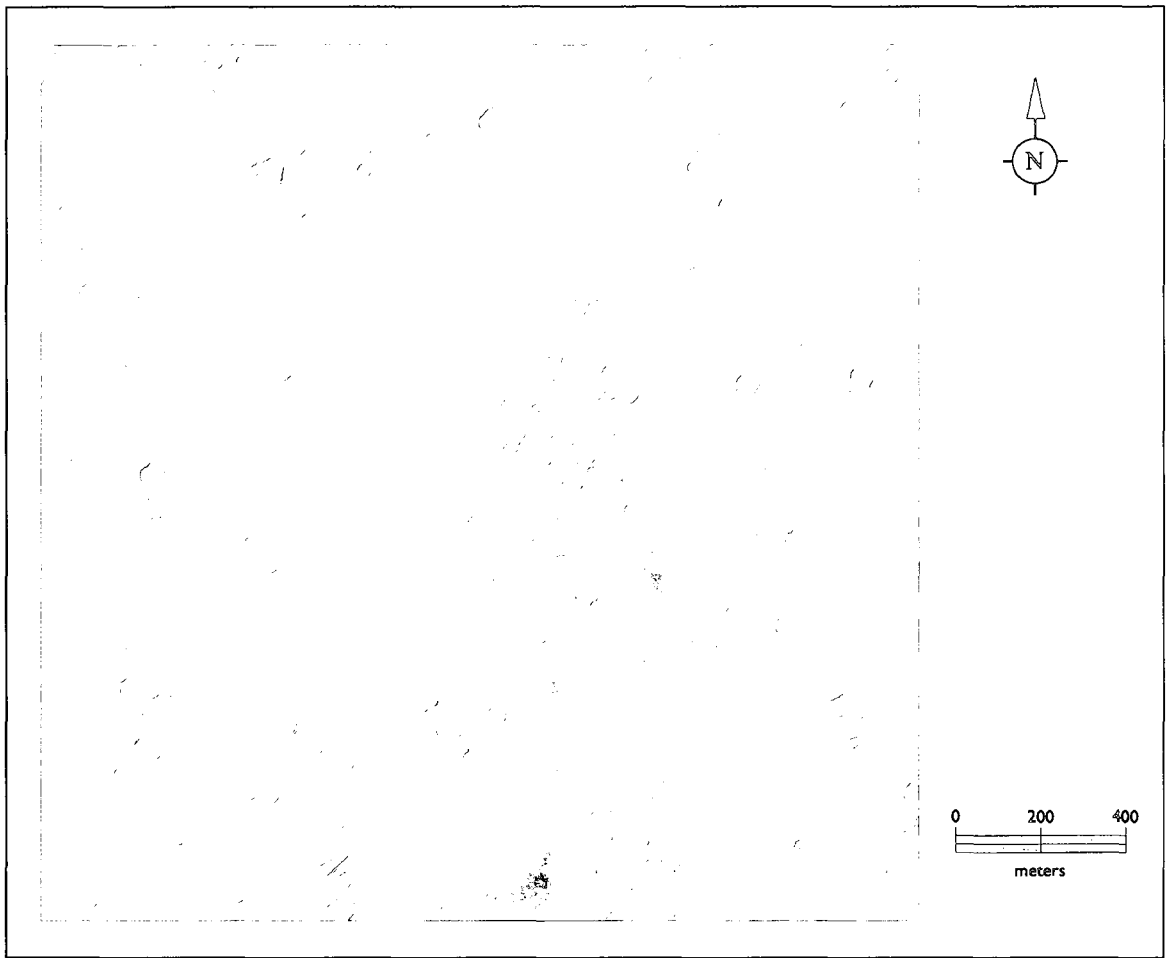
Figure 3. Columbia Hills site. Top: DRG, bottom: oblique view of DEM (DRG overlay).
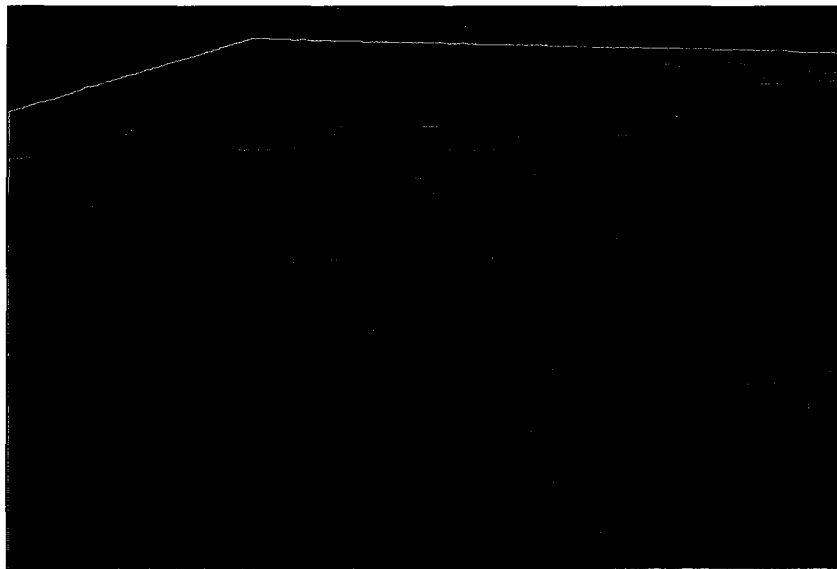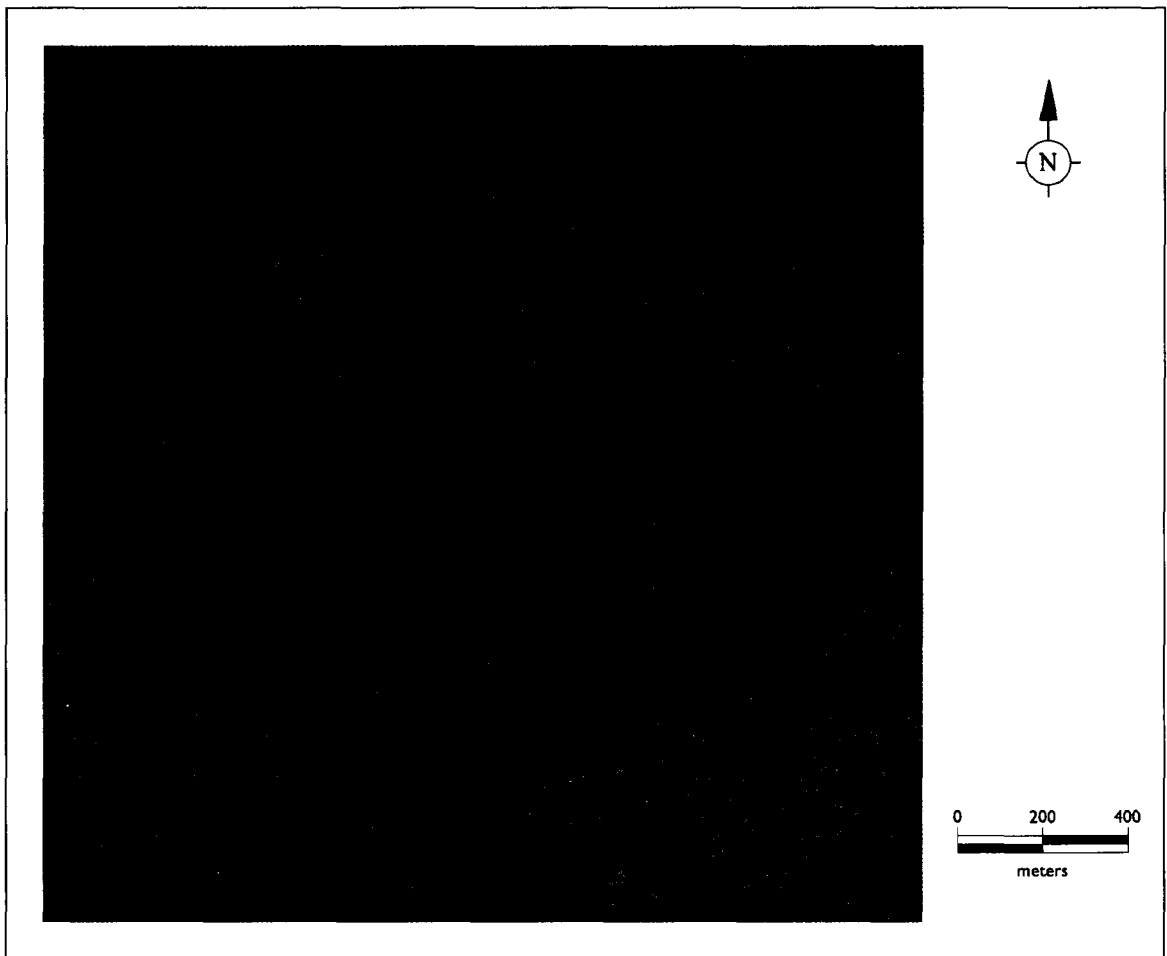
Figure 4. Gale Crater site. Top: DRG, bottom: oblique view of DEM (DRG overlay).
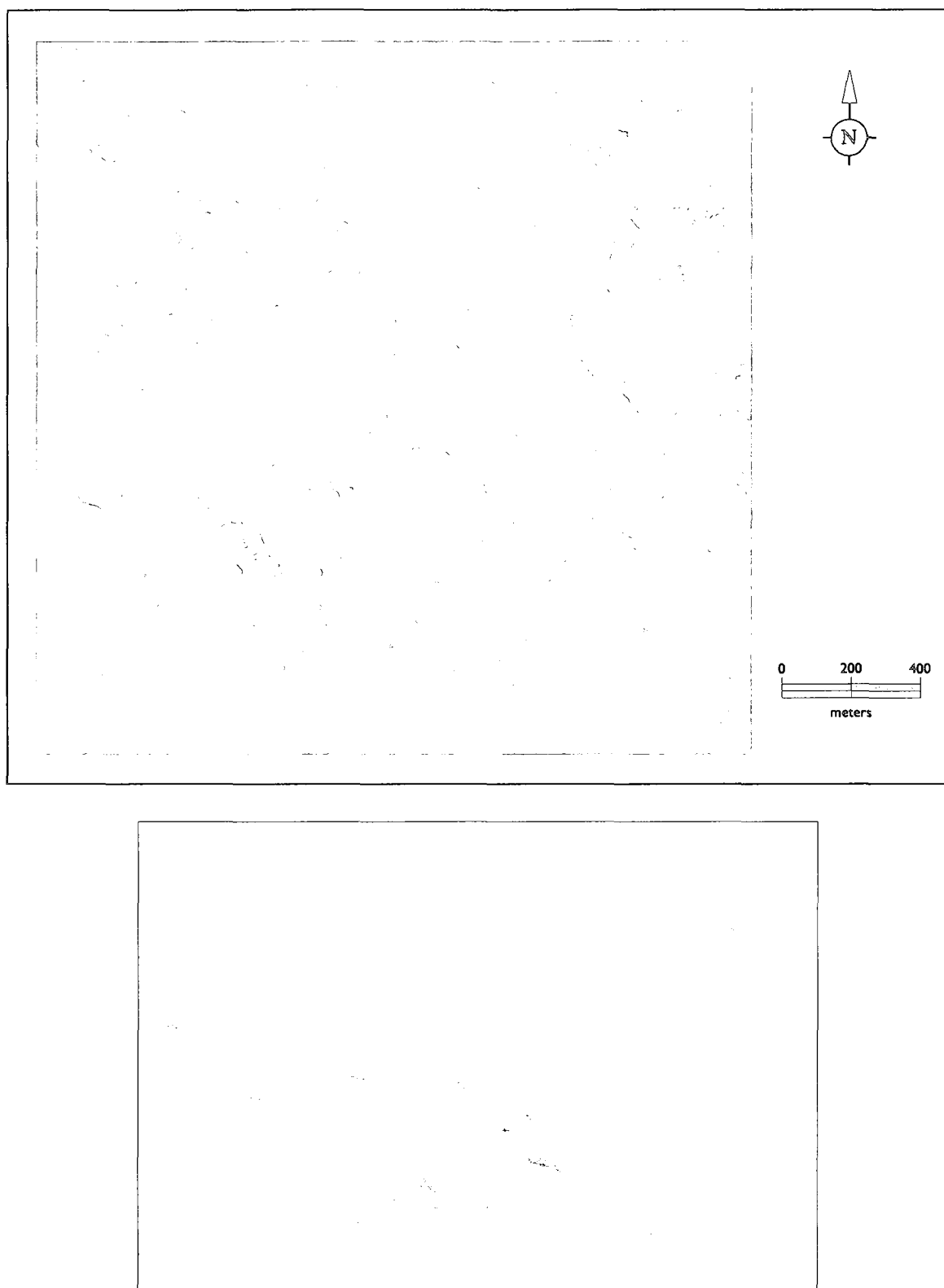
Figure 5. Mawrth Vallis site. Top: DRG, bottom: oblique view of DEM (DRG overlay).

CHAPTER 3: METHOD

Approach

This work followed the second trend in classification, that of self-selection, to

define the landform elements and used techniques from the classification-segmentation

group. A fuzzy K-means cluster analysis was used for classification, and segmentation

was performed by combining neighboring posts of similar clusters. Scale-space was used

to control high level of detail.

This approach eliminated the need for much of the specialized knowledge of the

MRO data sets that is currently uncertain or unavailable. A number of methods described

above employed a hydrological model based on terrestrial processes. Although recent

discoveries indicate the current and past presence of water on Mars, its influence in the

shaping of landforms is yet uncertain. To be as objective as possible, Earth-based

hydrological models and any other data that may introduce bias into the analysis were

avoided.

Hard limits for characteristics of the features to be extracted were required by

some methods. For example, the maximum size of the landform element or level of

separation between classes must be known and included in the procedure. These

parameters affected the outcome of the classification and were considered somewhat

arbitrary at this point. The fuzzy K-means method was chosen because it did not require

any such parameters.

Segmentation was performed on the results of the cluster analysis and was a

simple procedure based on Waldo Tobler's first law of geography "Everything is related

to everything else, but near things are more related than distant things." Based on this

law it was assumed that adjacent posts of the same class could be combined to create a

landform element.

The Columbia Hills site presented high detailed terrain. A sample of the site can

be seen in a picture from the Mars Exploration Rover, Spirit, showing many boulders and

small dunes (Figure 6) that could produce significant detail at a 1 m scale. The edge of

the rover's solar panels are visible in the foreground (the rover is about 1.5 m wide).
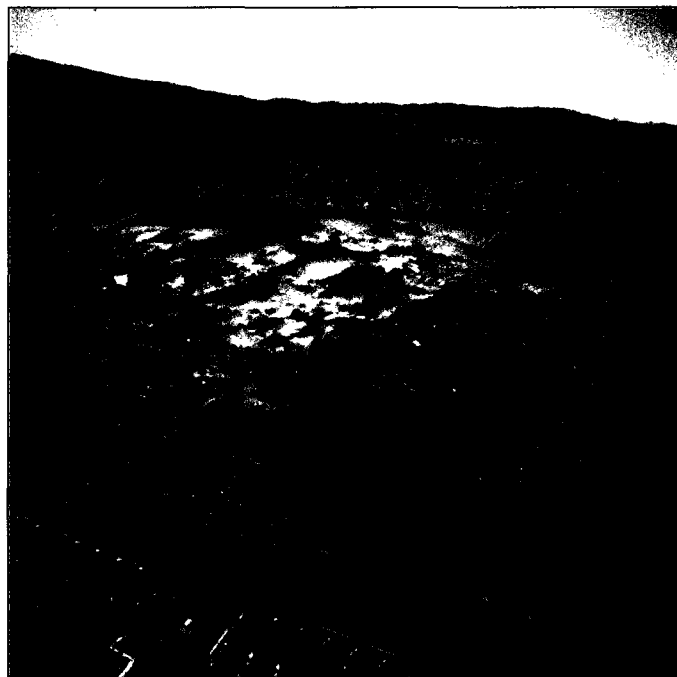
Figure 6. View of terrain in Columbia Hills from Spirit (courtesy JPL/NASA).

A 3D view of the Columbia Hills site with 2 times vertical exaggeration shows the irregular surface of the DEM (Figure 7). The computer can detect very small differences in elevation that the human eye may not notice. Even the seemingly easily discernible crater rings are broken and uneven. High frequency data such as this can produce over-segmented results unusable in this context. One remedy to reduce detail is to sub-sample the DEM. Several of the methods described earlier took this approach. This procedure causes arbitrary loss of data and was considered unacceptable. Other options included merging similar elements smaller than a given size to make new larger elements or using a smoothing function to remove the details and reveal the underlying structure of the data at a given scale.
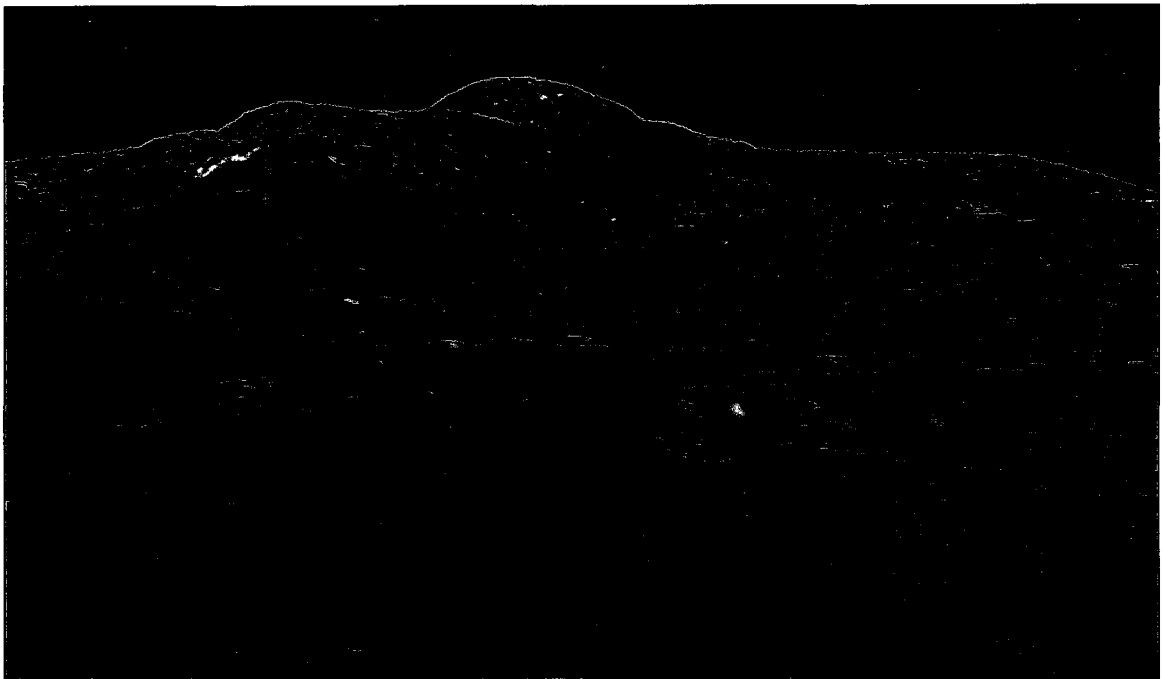
Figure 7. Oblique view of terrain in Columbia Hills DEM at 2X exaggeration.

In this work, detail was addressed through the use of a linear scale-space. Scale-space is a formal theory for handling image structures at multiple scales. A DEM is essentially an image of elevation. Instead of color, each pixel represents a height. A linear (Gaussian) scale-space uses a Gaussian kernel to smooth or blur objects in the image. The result is a coarser level of scale corresponding to the variance t of the kernel. As t increases, the resulting image becomes more coarse. Image structures smaller than approximately $\sqrt{t}$ are smoothed away at scale t. At the same time, prominent features and trends are preserved.
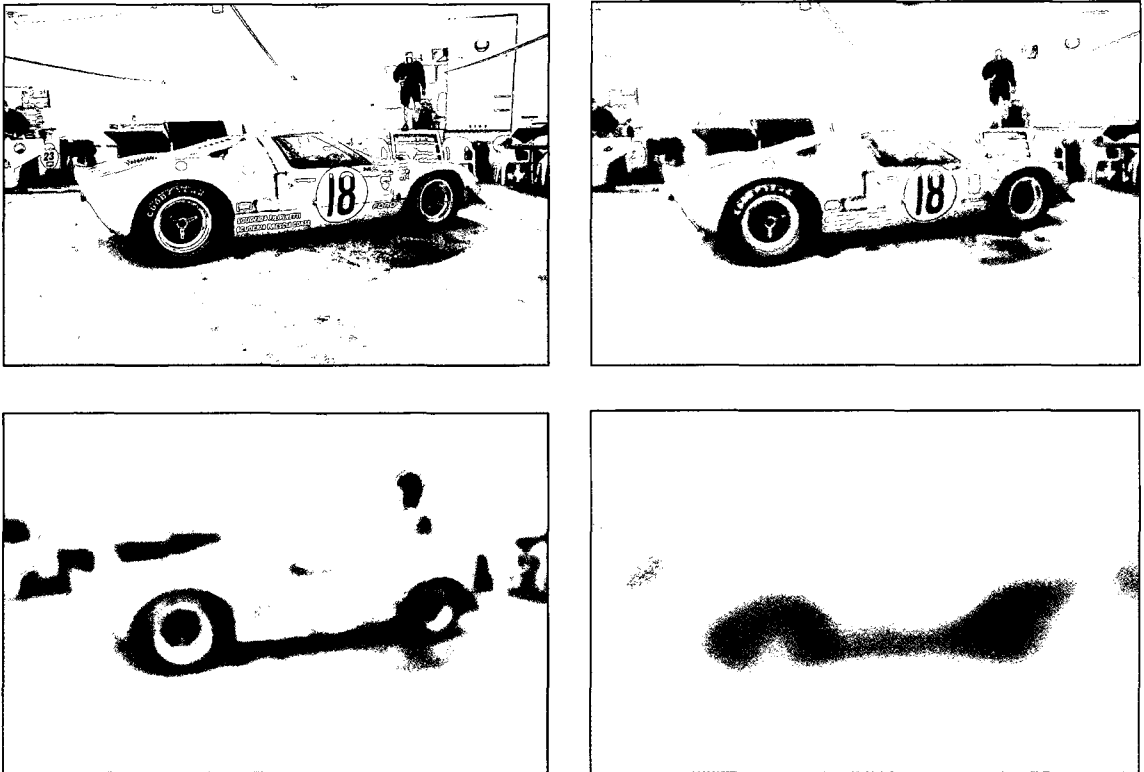
Figure 8. Different levels in the scale-space of an image at t = 0, 4, 16, & 64.

Figure 8 shows the results of a scale-space framework on a photo of a red Ford GT40 in its pit at the race track. The scale parameter t was applied at levels of 0, 4, 16, and 64, $t = 0$ representing the original photo. At $t = 4$, small details such as lettering and the objects on the tables and tool chests are blurred. At $t = 16$, larger details such as hubcabs and the mechanic's legs and arms disappear. At $t = 64$, only the major objects in the scene are visible. As t increases, the surface is smoothed, retaining the major components of the scene.

Several landforms stand out in the three study sites: dunes, hills and craters. Craters are important landforms in that they reveal the material underneath the surface to the observer. The craters in the Columbia Hills study site are relatively small by comparison to those examined in previous work (Bue & Stepinski, 2006; Stepinski et al., 2006, 2007) and are more difficult to differentiate from the high detail of the surrounding terrain. Reducing detail and thus increasing homogeneity in the crater rim should make it more discernible as a ring. The combination of the ring with the characteristic central depression should help to identify the crater. As a test of the usability of the newly-created landform elements, a search for craters was performed. This search entailed building objects from landform elements and testing them for crater characteristics.

In summary, fuzzy K-means cluster analysis was used for classification. It required very little data-specific knowledge and no hard limits on feature characteristics. Segmentation was a simple procedure performed on the results of the cluster analysis and based on similar adjacent posts. The MRO data are of a higher resolution than any to

date and over-segmentation may result in many small unusable elements. Scale-space was applied to the DEM to reduce over-segmentation at a minimal loss of data. Finally, landform elements were searched to find a common landform object, the crater.

Data Preparation

The DEM data sets were obtained from a USGS website (USGS, 2009) as large swaths (on the order of 10,000 x 20,000 posts per DEM). These files were stored in a standard format and projected. Their size placed a substantial burden on computer resources for the project. As such, a subsection of each study site was created at the size of 2048 x 2048 posts (approximately 2 km x 2 km). Many references applied a low pass filter to the raw data prior to processing to remove artifacts and noise. In tests, a low pass filter altered the outcome of the cluster analysis noticeably. To preserve as much information as possible, no filters were applied to the raw data.

The attributes which became components of the feature vector were chosen on the basis of their contribution to landscape characterization. This is the subjective part of the process and would nominally be based on expert opinion. For this project, attributes were selected from those commonly used in the literature including elevation, gradient, curvature, and relative elevation. Elevation derivatives were computed from the DEM

using a 3 x 3 matrix centered on each post and the partial quartic equation (Zevenbergen, 1987):

$$Z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Fxy + Gx + Hy + I$$

Elevation - This is the value obtained directly from the DEM. Elevation presented the widest numeric range of all of the attributes and was range-standardized to transform its values to lie between 0 and 100 inclusive.

Gradient - This attribute was computed as a first derivative of elevation, converted from degrees to a proportion, and multiplied by 100.

Curvature - Both plan and profile curvature were computed from elevation and then multiplied by 100 to give units of 1/100 LU. These values were in the same general range as elevation and gradient. The plan curvature is the rate of change in the aspect or surface normal and a negative or concave value represents a convergent flow down a slope. A positive or convex value indicates a divergent flow. Profile curvature is the rate of change in gradient. A negative or concave value represents a depression or valley. A positive or convex value indicates a mound or hill.

Relative Elevation - This attribute was not used in the fuzzy K-means procedure, but contributed later to the crater detection algorithm. The relative elevation of a given post was determined by finding the local elevation maximum and minimum in a given radius (20 posts) and then applying the following equation:

$$rel\_elev = 100(elev - local\_min)/(local\_max - local\_min)$$

A local depression is represented by a value of 0 and 100 represents a local peak. Figure 9 shows the relative elevation of the Gale Crater study site. The light areas (crater rims and dune tops) are high values and the dark areas (crater centers and bottoms of slipfaces) are low.

Figure 10 shows the histograms for each attribute of each study site. The elevation of both Columbia Hills and Mawrth Vallis is bimodal. The elevation of Columbia Hills and Gale Crater are both positively skewed. Additionally, the gradient data sets of all the study sites are positively skewed. All of the curvature histograms are exponential and also show extreme outliers. The locations of these values were examined in a 3D view and appear to be valid.
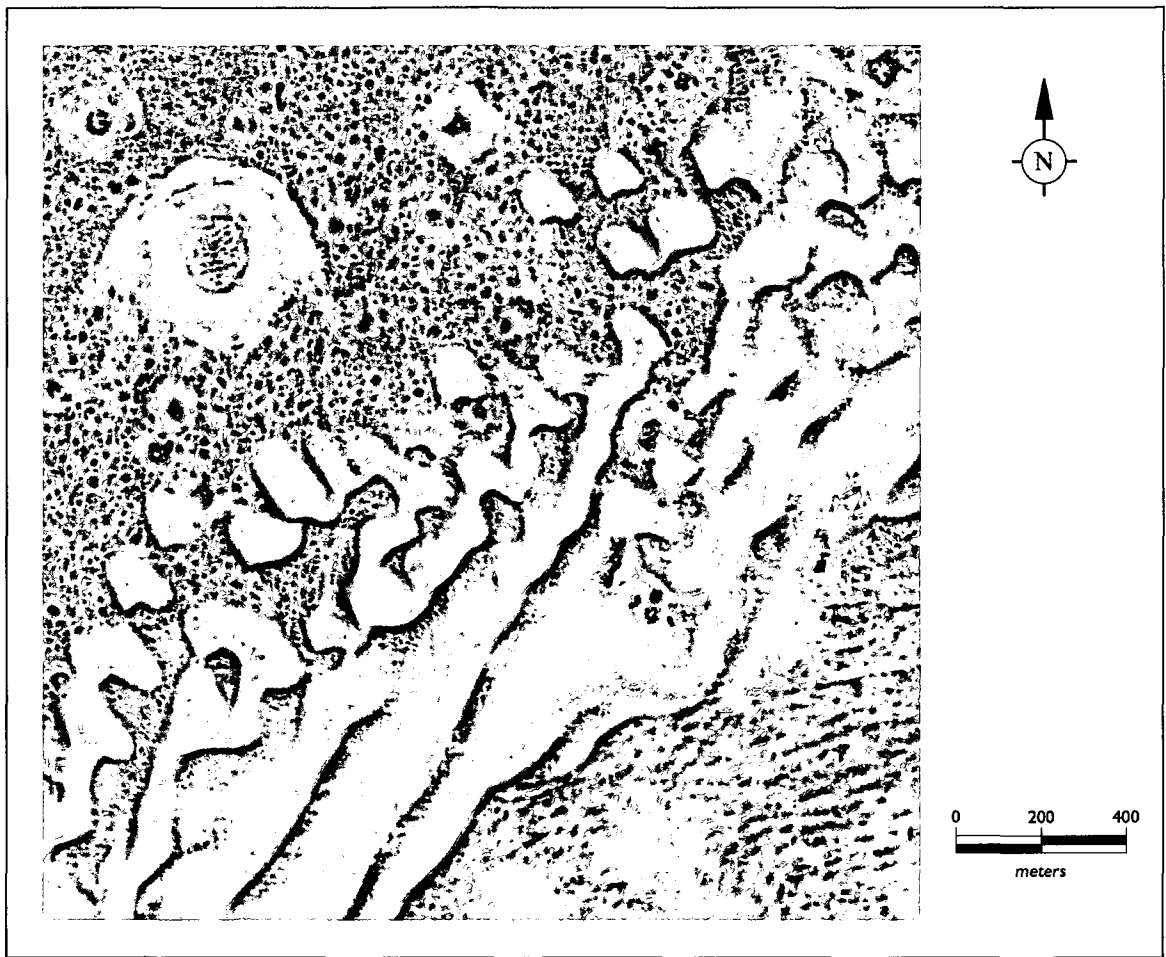
Figure 9. Relative Elevation of Gale Crater.

Elevation

| Columbia Hills | Gale Crater | Mawrth Vallis |

Relative Elevation

| Columbia Hills | Gale Crater | Mawrth Vallis |

Gradient

| Columbia Hills | Gale Crater | Mawrth Vallis |

Plan Curvature

| Columbia Hills | Gale Crater | Mawrth Vallis |

Profile Curvature

| Columbia Hills | Gale Crater | Mawrth Vallis |
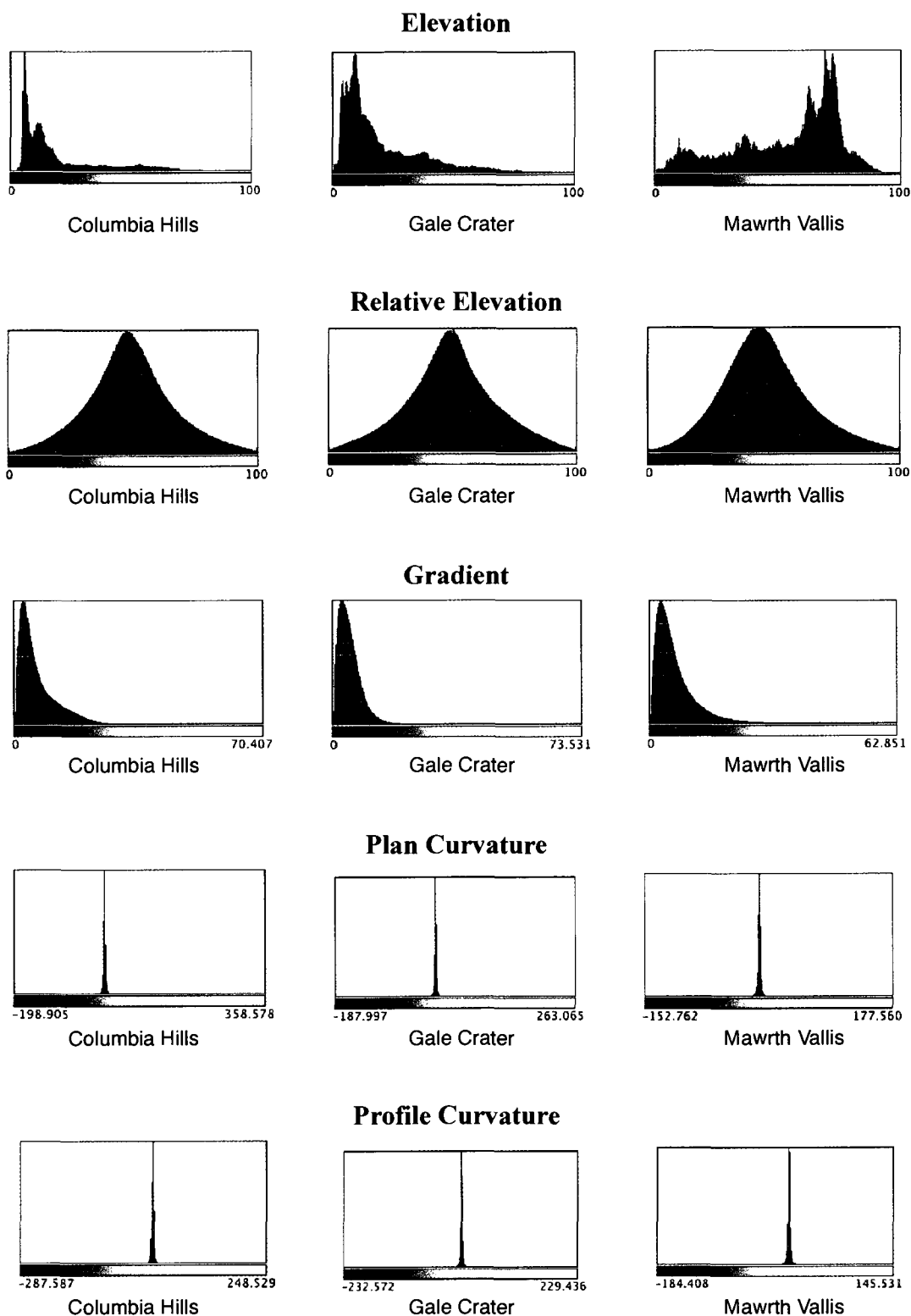
Figure 10. Histograms of study site data attributes.

30

Distance Metric

K-means employs a distance metric to measure the similarity of a post to a cluster and determine cluster membership. This metric is typically the Euclidean metric. Histograms (Figure 10) of the three study area feature vector component data sets show that elevation can be bimodal, gradient is skewed, and curvature is exponential. Together, these data produced a non-spherical distance distribution for which the Euclidean metric was not appropriate. Two other distance metrics were considered, Diagonal and Mahalanobis. The Diagonal metric compensates for the disparities in variances of the components of the feature vector. However, a covariance matrix (Table 2) of the Columbia Hills feature vector shows that some of the components have a negative covariance. The Mahalanobis distance metric, based on the data itself, was used to compensate for the non-spherical nature of the distribution and the negative covariance.

Table 2

Covariance Matrix of Columbia Hills Data Attributes

	Gradient	Plan Curvature	Profile Curvature	Elevation
Gradient	0.061292	0.000206	0.001240	-0.010618
Plan Curvature	0.000206	0.146786	0.059462	0.000307
Profile Curvature	0.001240	0.059462	0.123019	-0.000285
Elevation	-0.010618	0.000307	-0.000285	0.004548

Procedure

A Fuzzy K-means procedure using gradient, plan curvature, profile curvature, and relative elevation was developed for the ImageJ image processing software. ImageJ is open source software originally developed by the National Institutes of Health (NIH). Many others have contributed to ImageJ and extended its functionality through "plug-ins", software modules that can be "plugged into" an existing application. A fuzzy K-means plug-in was written for ImageJ and run for 2-10 clusters for each sample site, each post receiving a membership value for each cluster. The fuzziness factor was set at 1.5 (a typical value from the literature), the convergence threshold was set to 0.001, and the intragrade threshold was set at 0.6 (from Burrough et al., 2000). The cluster of highest membership was assigned to the post as its cluster number. Posts that did not meet the intragrade threshold were assigned the intragrade cluster number.

In the segmentation step, elements were formed by merging adjacent non-intragrade posts with the same cluster number, and each element was assigned an ID. Intragrades provided boundaries for objects of the same cluster and allowed some posts to remain unclassified. Each of the segmentations was saved as two images: the pixel value of the first consisted of the element ID, and the second, the cluster number. These images were used later to retrieve the results of the segmentation step for further analysis and visualization.

This procedure was applied to the raw DEMs and to DEMs within a scale-space. A Gaussian blur filter was applied to the raw data to produce DEMs at different scale-

space levels. The fuzzy K-means and segmentation procedures were then applied to these new data sets.

A Gaussian blur of scale parameter t ranging from 2 to 12 was applied to the Columbia Hills study site in the crater detection procedure. After the fuzzy K-means and segmentation procedures the cluster indicating the crater rim was identified by viewing the results. Another ImageJ plug-in was used to find occurrences of rim elements that contained a depression within their bounding box. The depression was determined by the relative elevation of the element. The depression was flooded to the lowest elevation of the rim element using the bounding box of the rim element to constrain the flood from overflowing in the case of a breached crater. Next, the compactness of the depression shape was tested for roundness using the following formula. A compactness value of > 0.75 was considered indicative of a possible crater.

$$compactness = 4\prod(area)/(perimeter)^2$$

Candidate craters were detected in each scale-space level and identified in a binary image where pixels with a value = 1 indicated the presence of a crater and a value = 0 indicated absence. The binary images for each scale-space level were combined with a binary OR operator. This combination represented the entire set of candidate craters found for the study site.

CHAPTER 4: RESULTS

To evaluate the fuzzy K-means results, the cohesion and separation values for the clusters were computed for each run. Cohesion is based on the distance of a post to its cluster center and separation is based on the distance from the overall center. The optimal cohesion/separation combination is a small value for cohesion and a large value for separation. The values for cohesion and separation for each sample site are shown in Figure 11. Along with cohesion and separation the effectiveness of the clusters must be examined, as well as inclusion of all the expected landform types, and cluster redundancy. A 3D view of the DEM overlaid with the results of the cluster analysis was invaluable in this process.

For each study site the classification improves with the increase in number of clusters. Analyses with cluster numbers of 5, 6, 7, and 8 are discussed further in this chapter. Figures 12 - 24 show analyses for each study area. Each color represents a cluster and color intensity represents membership or how strongly an object belongs to its cluster (brighter is stronger). Pixels that did not meet the intragrade threshold are colored black. Corresponding tables 4-19 show the range and mean for each component of the feature vector for each cluster and a related landform type. It is easy to see specific landforms in the results. Additionally, some linear artifacts from the camera and some introduced in the construction of the DEMs are visible.
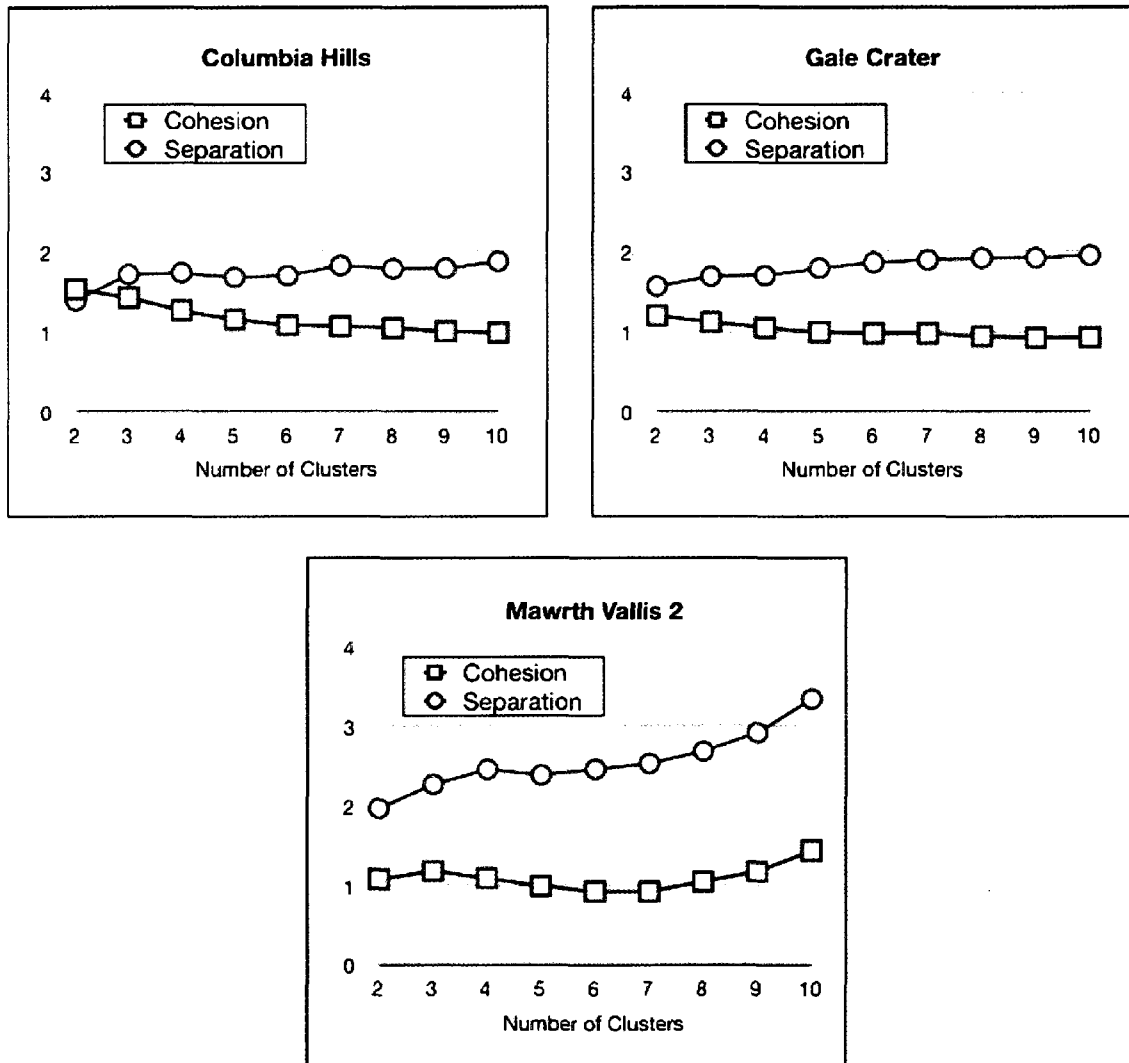
Figure 11. Graphs of separation and cohesion.

Each cluster was assigned a landform type according to the nomenclature used in Ventura and Irwin (2000) as a divergent or convergent shoulder, divergent or convergent backslope, divergent or convergent footslope, or a level slope. Shoulders have a positive profile curvature value and a convex form, whereas footslopes have a negative profile

curvature value and a concave form. Clusters with a zero profile (flat) are called

backslopes. Divergent forms have a convex plan curvature (positive value), whereas

convergent forms have a concave plan (negative value). Additional familiar landforms

that are identified later include dune windward slope, dune slipface, crater rim, crater

inner wall and crater ejecta. Table 3 shows the relation of plan and profile curvature to

landform types.

Table 3

Curvature-based Classification for Landform Elements

Plan Curvature	Profile Curvature	Landform
positive	positive	divergent shoulder
negative	positive	convergent shoulder
0	positive	shoulder
positive	negative	divergent footslope
negative	negative	convergent footslope
0	negative	footslope
positive	0	divergent backslope
negative	0	convergent backslope
0	0	backslope
0	0	level slope

Columbia Hills

The Columbia Hills study site can be described as an outcrop surrounded by a plain marked with many craters. Because of the high detailed and irregular nature of the plain, the cluster analysis has broken it into many tiny segments defined by changes in plan and profile curvature. The addition of more clusters (7 and 8) reduces this over-segmentation as the process assigns a cluster specifically to the gentle backslope that fills much of the terrain surrounding the hills. The crater walls are easily visible and the slopes of the outcrop are segmented based on elevation and gradient. Results for 5 and 6 clusters split the hills into an upper shoulder and a steeper lower footslope. Results for 7 and 8 clusters show the same footslope but the upper shoulder is split into a crest and a shoulder. The hill tops and ridges are much more visible. One can see the peaks of Brown, Chawla, Clark, and Husband, and the ridge east of Husband Hill (shown in Figure 14). In 6 clusters a low shoulder can be detected running from north to southeast in the eastern half of the site. This slope, although more detailed, is visible in the 7 and 8 cluster results.

Figure 12. Fuzzy K-means clustering of Columbia Hills site at 5 clusters.

Table 4

Columbia Hills attribute range and mean for 5 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,41] 12	[0,11] 4	[-0,15] 1.7	[-18,0] -1.9	divergent footslope
2	[0,97] 24	[9,48] 13.2	[-16,20] 0.1	[-18,15] -0.7	divergent footslope on hills, crater inner wall
3	[31,98] 45.8	[0,22] 8.9	[-10,8] -0.1	[-8,11] 0.1	convergent shoulder on hills
4	[0,46] 12	[0,12] 4	[-6,8] 0.4	[0,23] 2.1	divergent shoulder
5	[0,45] 11.9	[0,12] 4	[-21,-0] -2.2	[-8, 9] 0.1	convergent shoulder

38

Figure 13. Fuzzy K-means clustering of Columbia Hills site at 6 clusters.

Table 5

Columbia Hills attribute range and mean for 6 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[31,96] 46.8	[0,21] 8.6	[-10,10] 0.0	[-9,13] -0.1	shoulder on hills
2	[0,46] 11.6	[0,12] 3.5	[0,19] 2.2	[-8,9] 0.1	divergent shoulder
3	[0,42] 11.8	[0,10] 3.5	[-15,0] -1.8	[-0,16] 1.8	convergent shoulder
4	[19,97] 43.3	[11,43] 16.1	[-12,12] -0.1	[-16,11] -0.7	convergent footslope on hills
5	[0,45] 11.8	[0,11] 3.7	[-8,7] -0.3	[-21,-0] -2.2	convergent footslope
6	[0,37] 12.7	[5,16] 8.1	[-7,6] 0.0	[-6,9] 0.4	crater inner wall, shoulder
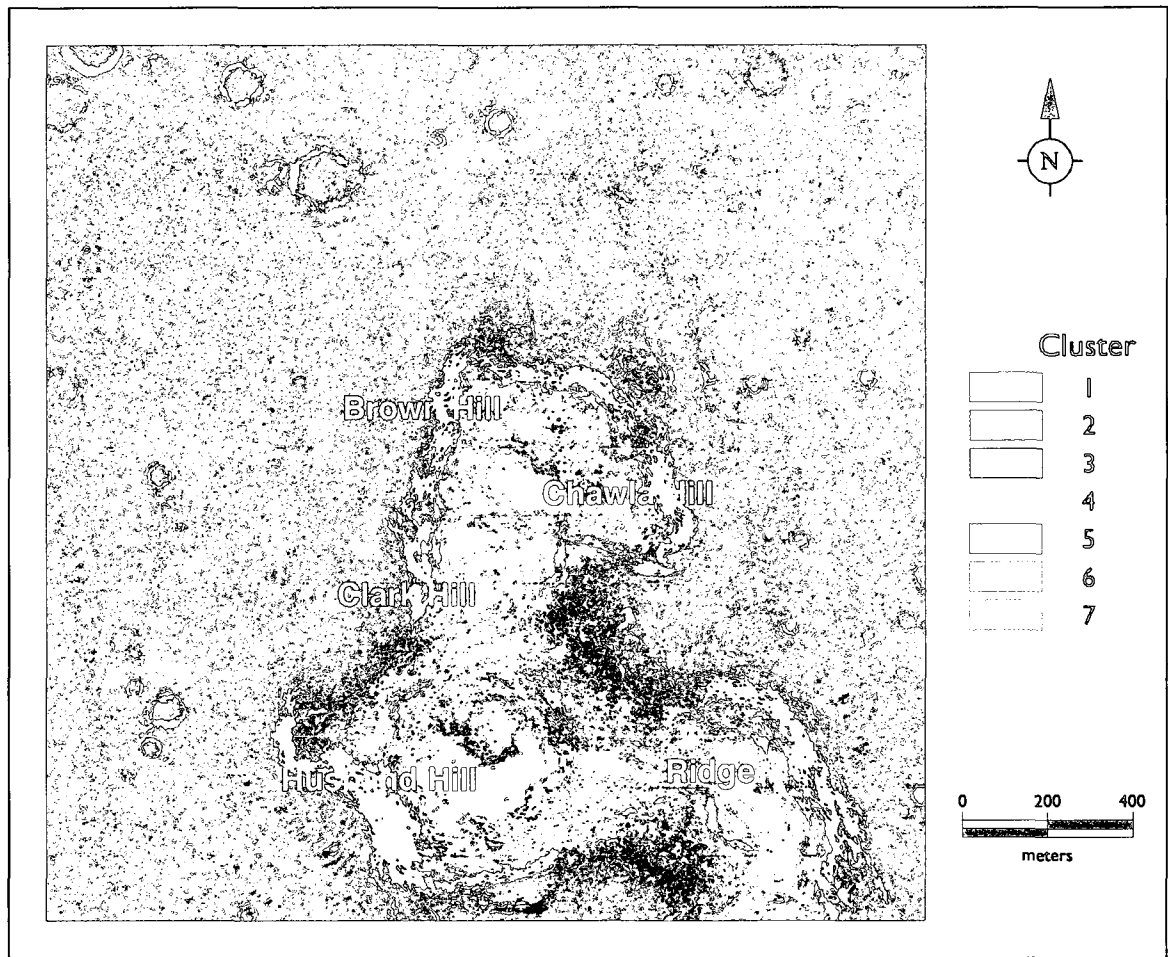
Figure 14. Fuzzy K-means clustering of Columbia Hills site at 7 clusters.

Table 6

Columbia Hills attribute range and mean for 7 clusters

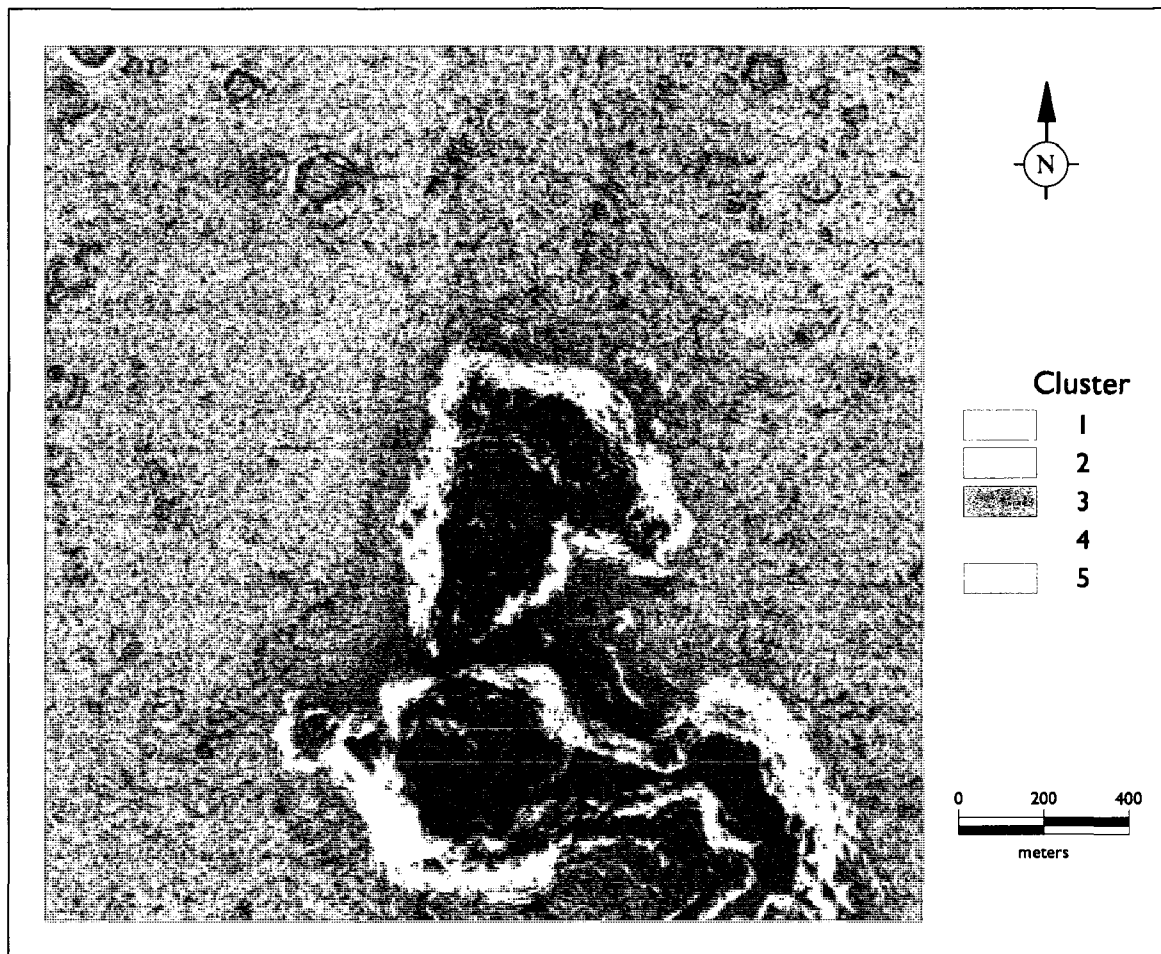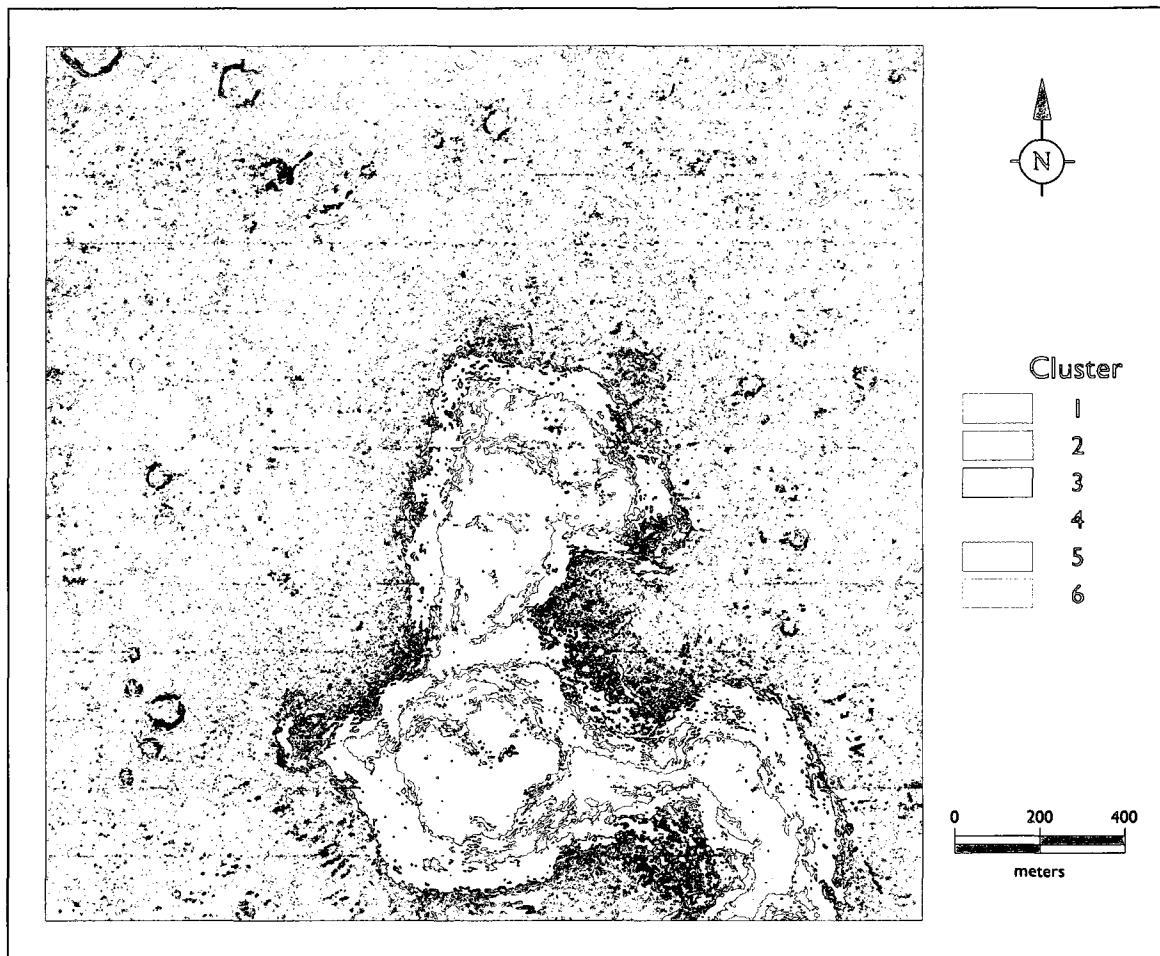| Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,31] 11.1	[4,13] 7.1	[-4,8] 0.5	[-5,7] 0.0	divergent backslope
2	[32,98] 47.3	[7,26] 12.6	[-8,9] 0.0	[-10,10] 0.1	shoulder on hills
3	[0,40] 11.6	[0,10] 3.4	[-1,11] 1	[-19,0] -2.2	divergent footslope
4	[1,96] 28.7	[12,48] 16.6	[-16,13] -0.6	[-15,15] -0.8	convergent footslope on hills, inner crater wall
5	[31,97] 50.1	[0,19] 7.2	[-9,14] -0.2	[-10,11] 0.0	convergent backslope
6	[0,41] 11.3	[0,9] 3.1	[-3,11] 1.1	[-1,14] 1.5	divergent shoulder
7	[0,45] 11.6	[0,12] 3.5	[-22,-0] -2.6	[-4,13] 0.8	convergent shoulder

40
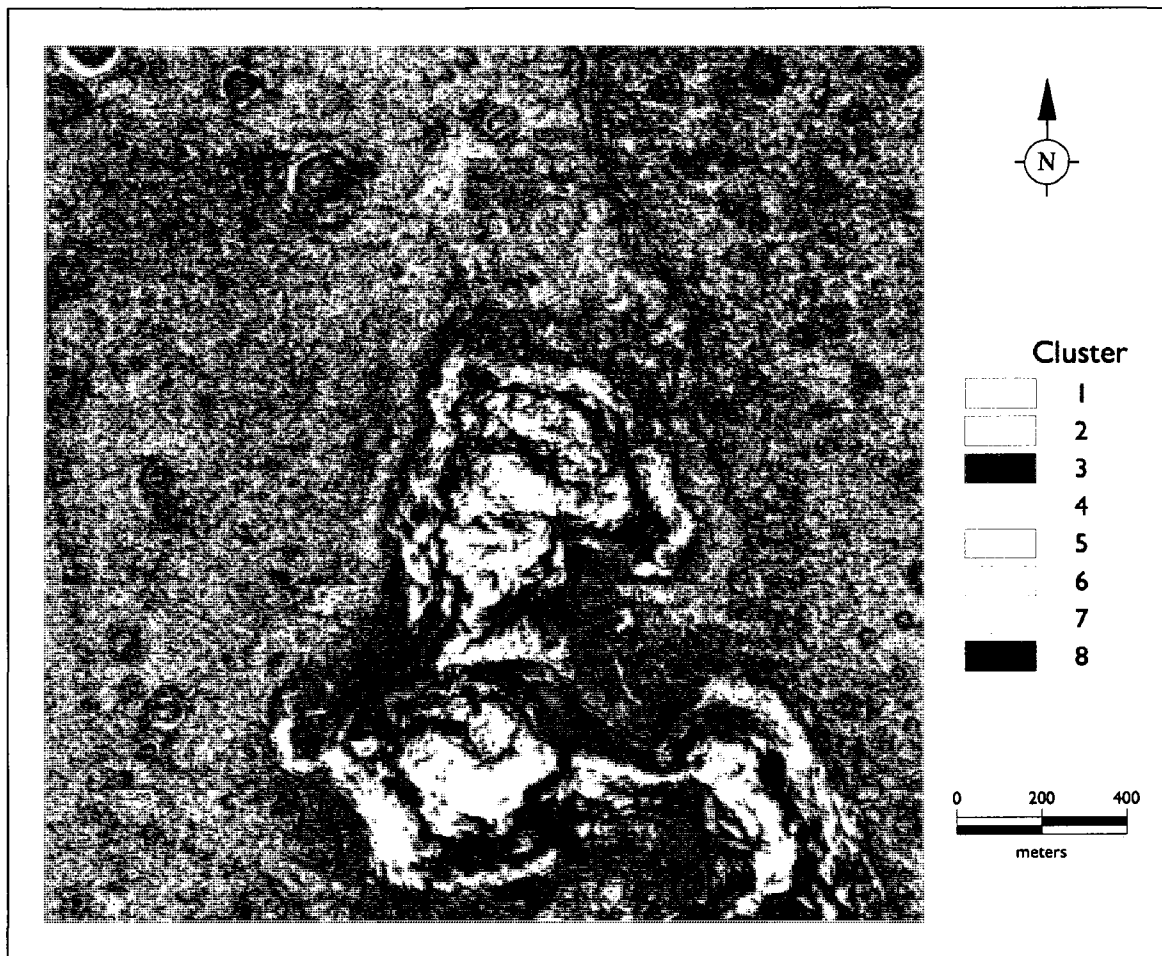
Figure 15. Fuzzy K-means clustering of Columbia Hills site at 8 clusters.

Table 7

Columbia Hills attribute range and mean for 8 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,48] 11.5	[0,14] 4	[1,22] 3.3	[-10,4] -0.7	divergent footslope
2	[0,46] 11.9	[0,15] 5.1	[-9,3] -0.7	[0,24] 3.3	convergent shoulder
3	[28,98] 44.1	[7,28] 12.3	[-8,8] 0.0	[-9,10] 0.0	slope on hills
4	[0,32] 11.2	[0,8] 3.3	[-2,2] 0.0	[-3,3] 0.1	backslope
5	[0,52] 11.7	[0,15] 4	[-23,-1] -3.3	[-4,11] 0.7	convergent shoulder
6	[33,97] 51.5	[0,19] 7.6	[-9,10] -0.2	[-10,11] 0.0	convergent backslope
7	[0,96] 23.6	[10,48] 14.9	[-16,13] -0.1	[-15,15] -0.4	convergent footslp on hills
8	[0,44] 11.7	[0,14] 5	[-3,8] 0.7	[-22,0] -3.1	divergent footslope

Gale Crater

The Gale Crater study site is an area of dunes that is sloping upward to the southeast with a large crater in the northwest. The lowland is also marked with many small craters or pits. The cluster analysis separated the site into 3 regions, each at a different level. These regions are designated A, B, and C and are clearly visible in all of the figures. A is blue in 5 and 7 clusters, green in 6 clusters, and yellow in 8 clusters. B is a divergent shoulder in 5 clusters, becomes a slope in 6 and 7, and splits into two clusters in 8. C is magenta in 5 and 8 clusters, cyan in 6 clusters and orange in 7 clusters. It changes from a divergent shoulder to a relatively flat one. Clusters 2 and 4 (green and yellow) display opposite but strong plan curvature with a concave profile curvature. These clusters are representative of the pitted terrain of A. The dune slipfaces and inner crater walls are visible as red in the 5-cluster analysis. The dune windward slope and crater outer wall (magenta and blue) become more apparent in 6 clusters. The slip faces and windward slopes become more distinct in the 7-cluster result and even more in the 8-cluster.

Figure 16. Fuzzy K-means clustering of Gale Crater Site at 5 clusters.

Table 8

Gale Crater attribute range and mean for 5 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,67] 16.1	[7,49] 10.3	[-16,11] -1	[-19,19] 0.1	slip face, crater inner wall
2	[0,29] 11.2	[0,9] 3.9	[-1,15] 2	[-11,11] -0.1	divergent footslope in A
3	[19,46] 26.7	[0,11] 5.7	[-6,4] -0.5	[-8,8] 0.1	convergent shoulder in B
4	[0,24] 12.3	[0,6] 4.2	[-10,2] -1.4	[-9,10] -0.2	convergent footslope in A
5	[43,71] 45.6	[0,17] 3.7	[-3,4] 0.6	[-4,4] -0.9	divergent footslope in C

Figure 17. Fuzzy K-means clustering of Gale Crater Site at 6 clusters.

Table 9

Gale Crater attribute range and mean for 6 clusters

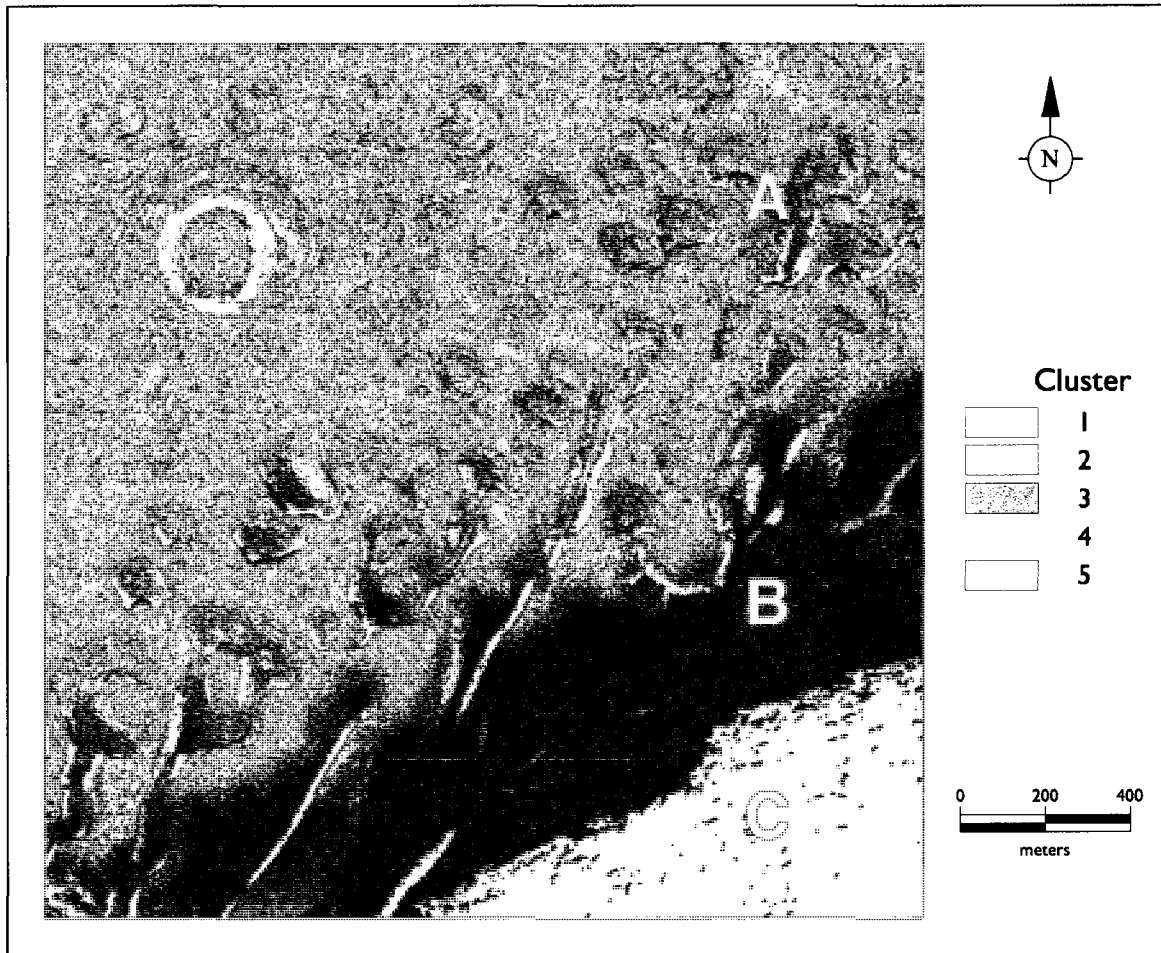| Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[1,75] 18.6	[8,51] 12.0	[-12,11] -0.2	[-20,19] 0.0	slip face, crater inner wall
2	[22,48] 29.1	[1,10] 5.3	[-4,4] 0.0	[-9,8] 0.0	slope in B
3	[0,35] 12.8	[1,10] 5.5	[-1,14] 2.2	[-11,11] -0.2	divergent footslope in A
4	[0,24] 11.6	[0,5] 3.1	[-3,4] 0.0	[-8,8] 0.0	slope in A
5	[0,36] 12.7	[0,10] 5.1	[-15,1] -2.3	[-10,11] 0.2	convergent shoulder in A
6	[49,72] 50.6	[0,16] 5.1	[-3,3] 0.1	[-5,6] -0.2	divergent footslope in C

Figure 18. Fuzzy K-means clustering of Gale Crater Site at 7 clusters.

Table 10

Gale Crater attribute range and mean for 7 clusters

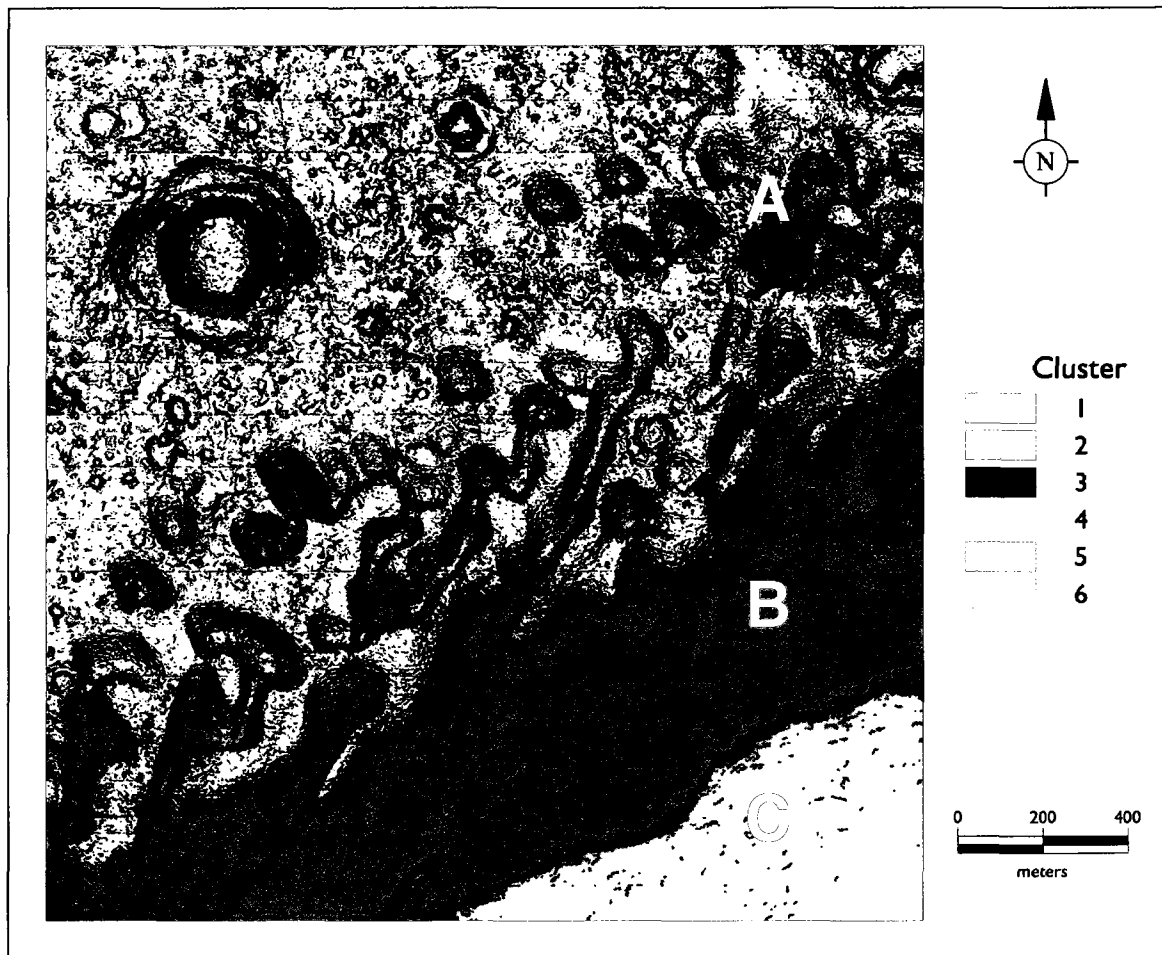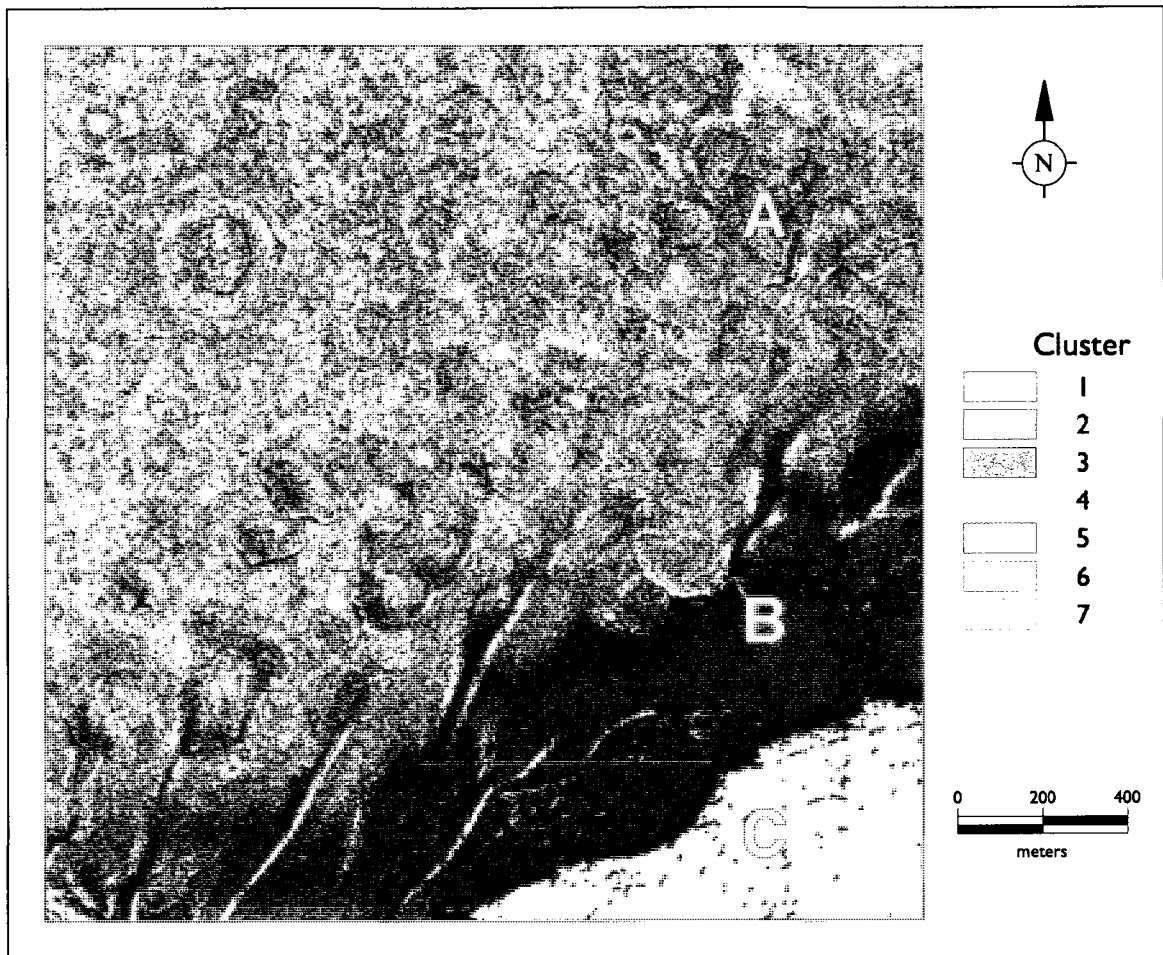| Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,24] 11.3	[0,5] 3.3	[-3,3] 0.0	[-7,7] 0.0	slope in A
2	[0,34] 13.9	[5,10] 7.1	[-3,3] 0.1	[-8,7] -0.3	windward dune slope, crater ejecta
3	[23,49] 29.4	[0,10] 4.9	[-4,4] 0.0	[-9,8] 0.0	slope in B
4	[0,36] 12.2	[0,10] 4.3	[-0,16] 2.6	[-11,10] -0.2	divergent footslope in A
5	[0,40] 12.8	[0,10] 4.4	[-17,0] -2.8	[-10,13] 0.5	convergent shoulder in A
6	[1,85] 20.7	[10,49] 14.6	[-15,13] -0.3	[-18,19] 0.3	crater inner wall,slipface
7	[51,72] 52.5	[2,15] 5.7	[-4,3] 0.2	[-4,7] -0.3	divergent footslope in C

45
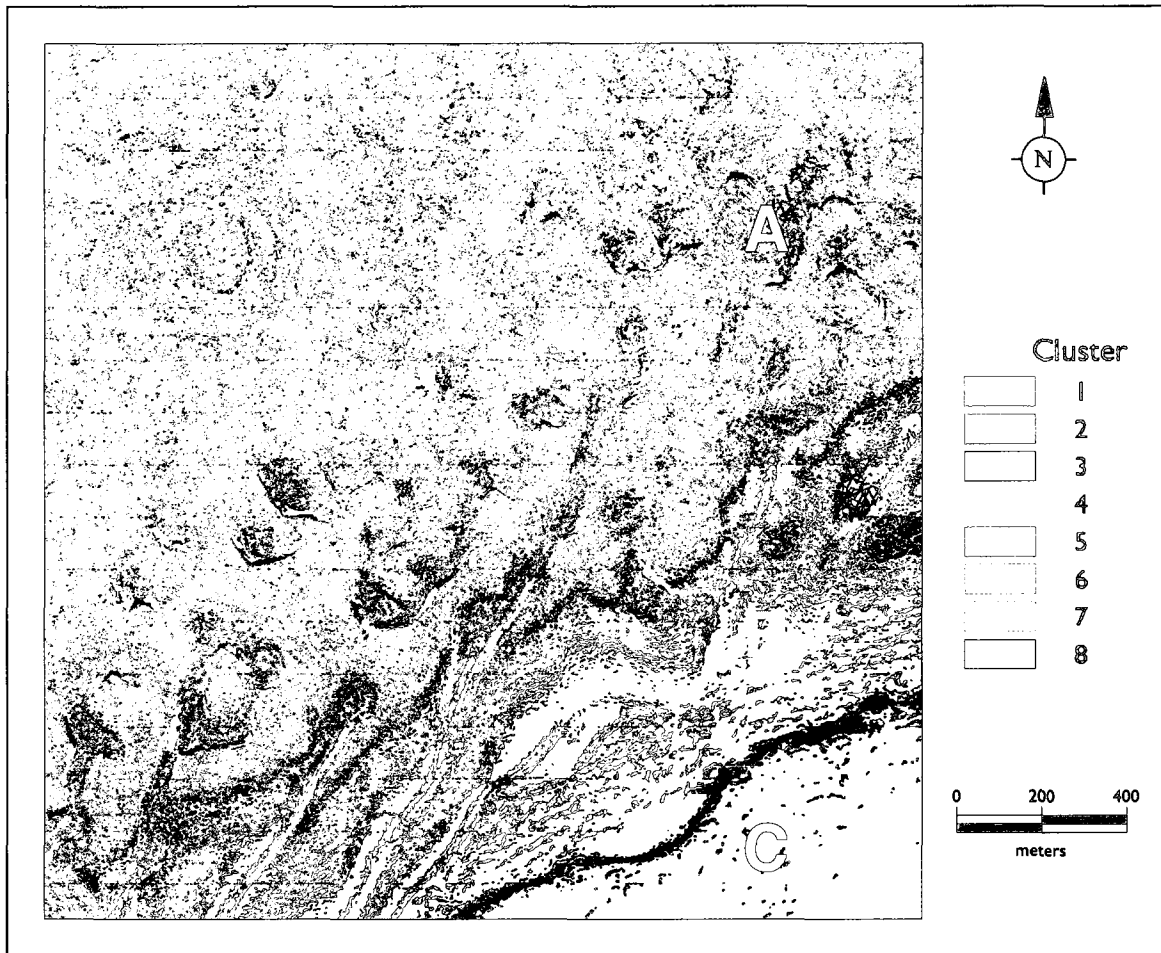
Figure 19. Fuzzy K-means clustering of Gale Crater site at 8 clusters.

Table 11

Gale Crater attribute range and mean for 8 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[0,23] 11.2	[0,5] 3.1	[-2,2] 0.0	[-7,6] 0.0	slope in A
2	[0,21] 10.7	[5,10] 6.6	[-3,4] 0.2	[-9,7] -0.4	windward slope,crater ejecta
3	[1,85] 19.1	[10,53] 14.8	[-16,15] -0.3	[-20,20] 0.3	inner crater wall, slip face
4	[25,52] 31.6	[0,7] 3.9	[-6,4] 0.0	[-8,10] 0.0	slope in B
5	[52,72] 54.3	[1,15] 6.7	[-6,6] 0.0	[-6,6] -0.1	footslope in C
6	[22,55] 31.2	[4,13] 7.4	[-5,5] -0.1	[-9,9] 0.1	convergent shoulder in B
7	[0,39] 12.4	[0,10] 4.1	[-17,-0] -2.8	[-10,13] 0.5	convergent shoulder in A
8	[0,35] 11.7	[0,10] 3.8	[-0,16] 2.5	[-11,10] -0.1	divergent footslope in A

46

Mawrth Vallis

The Mawrth Vallis study site is somewhat rough, with terrain that slopes downward from west to east. The terrain has sharp points and steep slopes. The cluster analysis separated the site into three regions which are designated as A, B, and C. Region C is colored blue in the 5 and 7-cluster images and cyan in the 6 and 8-cluster images. Region B is colored red, magenta, yellow, and blue in the 5, 6, 7, and 8-cluster images respectively. There is a relatively steeply-sloped region between in the upper middle part of area B designated area D. The type of steep slope in area D is also found throughout region A as hills. Some of the more level magenta is also shown on hill tops, and footslopes for the hills are shown in yellow. The 6 cluster result separates the steep slopes to those in the middle level of region B and those of the higher region A. The 7-cluster result adds a shoulder below the footslopes of the hills in region A. The 8-cluster result shows an interesting layering effect in the steeper hills of region A (Figure 20). The layers alternate between a convergent footslope and a divergent shoulder.
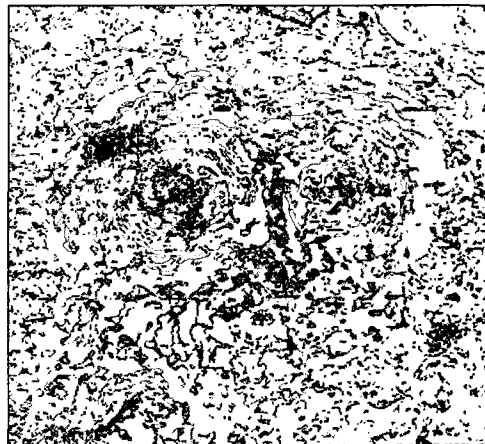
Figure 20. Layering in Mawrth Vallis site.

Figure 21. Fuzzy K-means clustering of Mawrth Vallis site at 5 clusters.

Table 12

Mawrth Vallis attribute range and mean for 5 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[32,58] 49.7	[0,11] 6	[-5,5] -0.2	[-5,7] 0.2	convergent shoulder in B
2	[5,99] 63.7	[12,53] 19.9	[-18,17] -0.4	[-23,22] -0.2	crest, ridge, region D
3	[12,31] 28.6	[1,14] 6.9	[-6,3] -0.2	[-6,8] 0.1	convergent shoulder in C
4	[48,100] 68.2	[6,15] 9.3	[-7,7] -0.2	[-10,9] -0.2	convergent footslope in A
5	[53,99] 68.6	[0,6] 4.3	[-7,6] -0.4	[-9,10] 0.3	convergent shoulder in A

Figure 22. Fuzzy K-means clustering of Mawrth Vallis site at 6 clusters.

Table 13

Mawrth Vallis attribute range and mean for 6 clusters

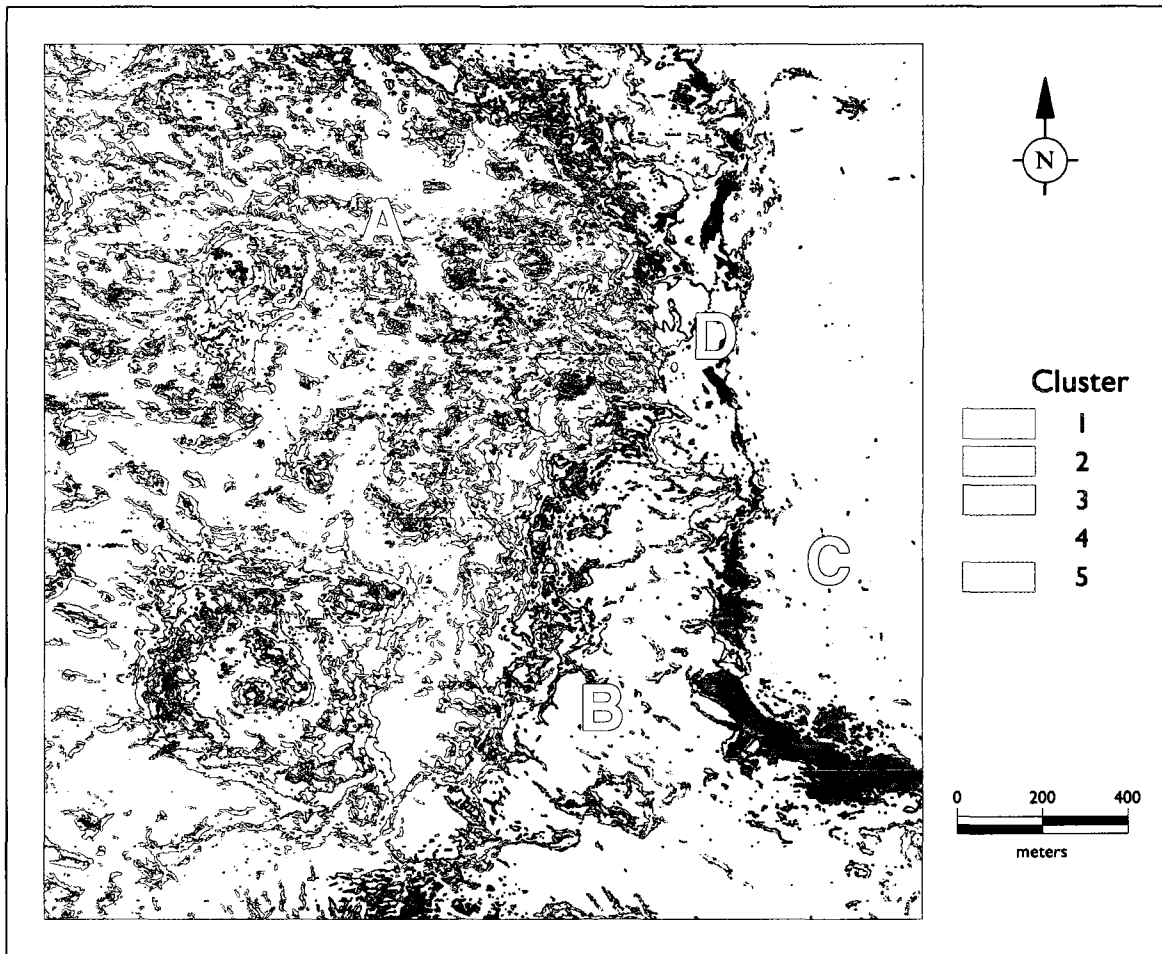| Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[54,92] 69	[0,5] 3.7	[-7,5] -0.5	[-7,8] 0.3	convergent shoulder in A
2	[4,63] 43.6	[9,24] 12.5	[-8,9] 0.3	[-10,12] 0.5	region D
3	[37,99] 67.2	[13,52] 19.9	[-15,10] -1.3	[-19,13] -1.7	crest, ridge
4	[53,100] 69.2	[6,13] 8.1	[-6,6] 0.0	[-8,8] 0.0	slope in A
5	[30,57] 47.9	[0,9] 5.3	[-5,4] -0.3	[-5,7] 0.1	convergent shoulder in B
6	[4,31] 27.2	[1,12] 6.5	[-6,3] -0.3	[-6,6] 0.1	convergent shoulder in C

49

Figure 23. Fuzzy K-means clustering of Mawrth Vallis site at 7 clusters.

Table 14

Mawrth Vallis attribute range and mean for 7 clusters

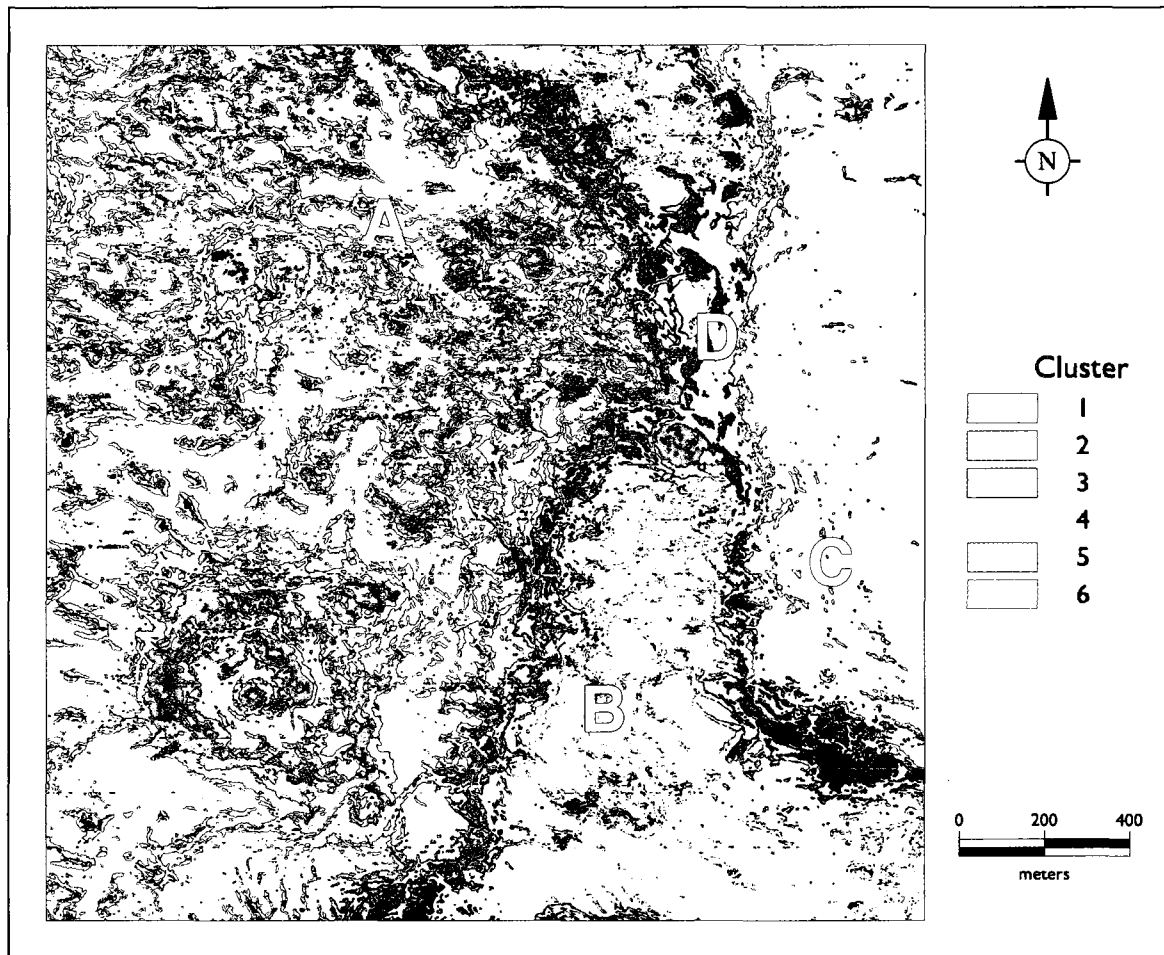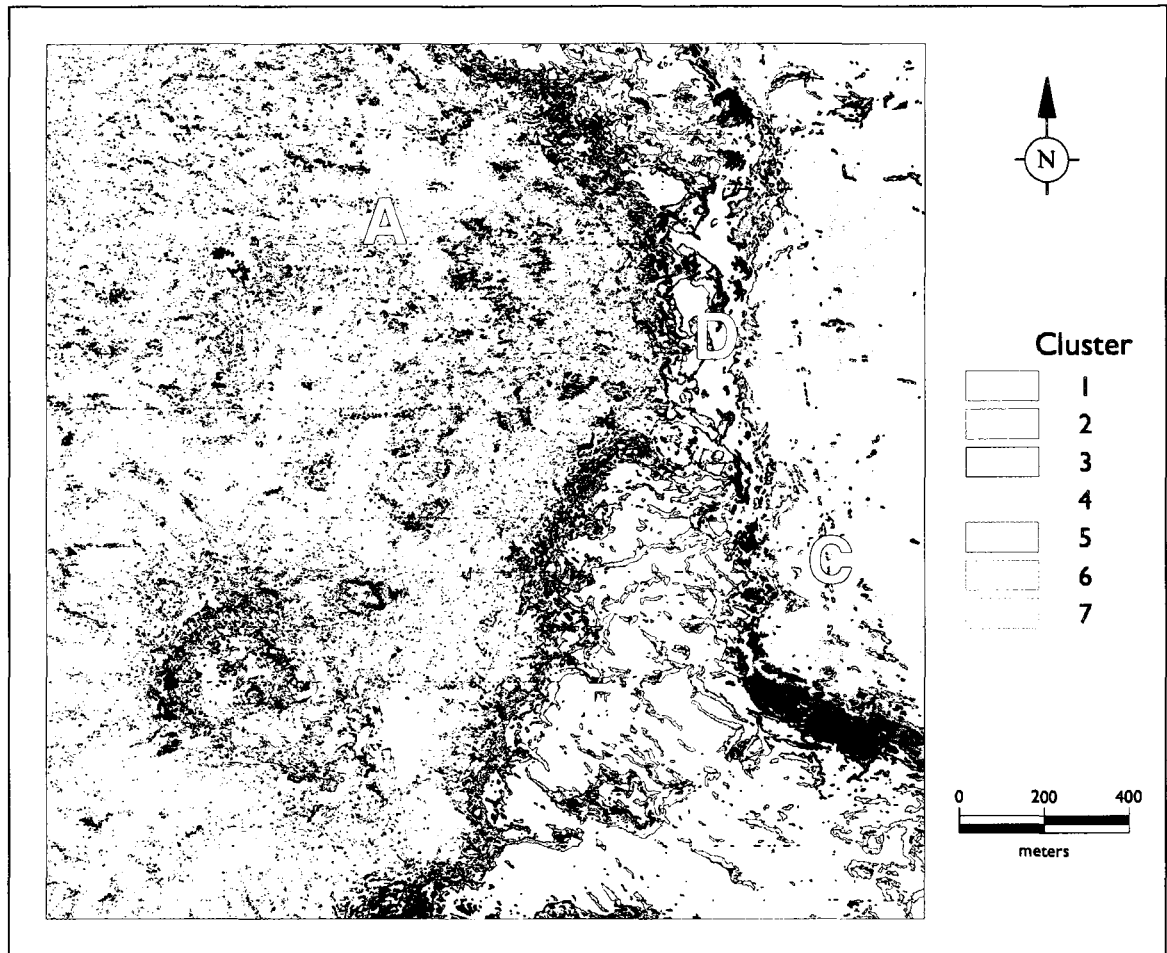| Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[25,99] 66.1	[13,52] 21.1	[-17,15] -0.2	[-18,20] 1.3	crest, ridge
2	[45,100] 68.3	[5,17] 9.5	[-12,6] -1.2	[-17,6] -2.2	convergent footslope in A
3	[4,30] 26.2	[1,11] 6.6	[-6,3] -0.3	[-6,5] 0.2	convergent shoulder in C
4	[29,56] 48.4	[0,8] 5.4	[-5,4] -0.4	[-6,7] 0.0	convergent backslope in B
5	[53,99] 69.4	[0,6] 3.9	[-8,4] -0.9	[-7,6] -0.4	convergent footslope in A
6	[50,100] 69	[2,13] 6.6	[-4,10] 1	[-5,12] 1.3	divergent shoulder in A
7	[3,61] 41.9	[8,21] 11.9	[-7,7] 0.2	[-10,10] 0.1	region D
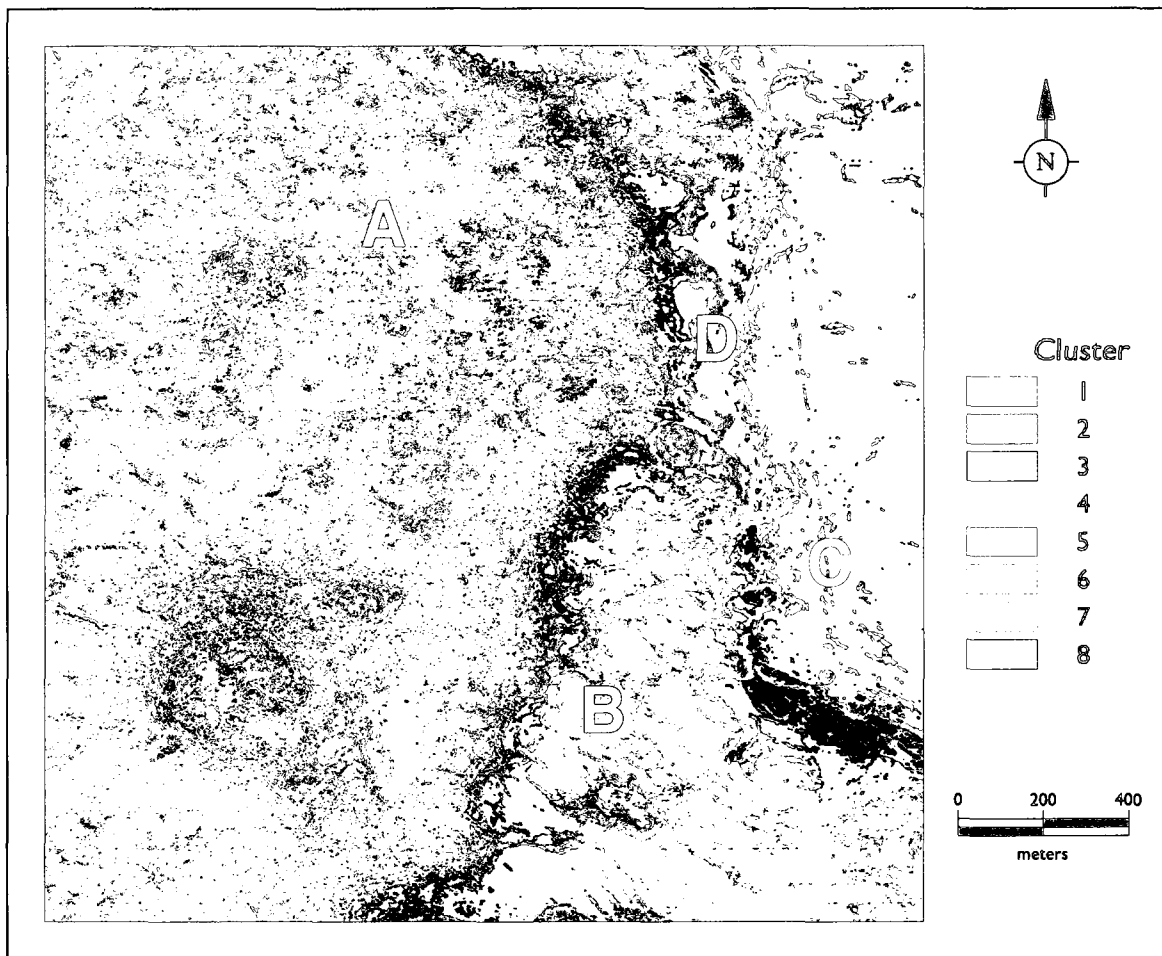
Figure 24. Fuzzy K-means clustering of Mawrth Vallis site at 8 clusters.

Table 15

Mawrth Vallis attribute range and mean for 8 clusters

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[53,99] 69.3	[0,6] 3.8	[-8,3] -1	[-9,5] -0.4	convergent footslope in A
2	[49,100] 68.8	[6,15] 8.9	[-9,5] -0.8	[-9,6] -1.1	convergent footslope in A
3	[28,56] 47.4	[0,8] 5.1	[-5,4] -0.4	[-6,7] 0.1	convergent shoulder in B
4	[5,99] 63.4	[7,56] 18.3	[-26,15] -1.9	[-42,5] -7.3	convergent footslope in A
5	[51,100] 68.8	[0,10] 4.9	[-3,9] 1.2	[-5,10] 0.9	divergent shoulder in A
6	[4,29] 25.2	[1,11] 6.4	[-6,3] -0.3	[-3,6] 0.2	convergent footslope in C
7	[23,99] 65.5	[10,53] 18	[-15,17] 0.3	[-6,31] 5.4	divergent shoulder in A
8	[2,58] 39.6	[8,24] 12	[-8,8] 0.2	[-8,7] -0.2	region D

Each study site was preprocessed with a gaussian blur of $2 < t < 12$. The value of $t = 5$ removes details that are < 2.24 m in diameter. Figures 25-27 show the results of the fuzzy K-means analysis. Associated tables 16-18 provide the range and mean for the feature vectors. For the Columbia Hills site this process removed enough of the detail in the crater rims to allow the cluster analysis to identify these landforms. The crests and ridges are clearly viewable as well (green). The steep lower slopes of the hills and the footslopes between them are also identified (red and cyan, respectively).

In the Gale Crater site, the dunes were separated into a footslope (blue) which indicates the gentler gradient at the bottom of the slipface and at the windward slope, a higher level shoulder (cyan) at the top and side of the slipface, and the top of the dune (red). These elements could be combined to define a single dune.

The Mawrth Vallis site clearly shows hill footslopes as green, steep upper slopes as red, and the crests and ridges as cyan. The gentle slopes of each level are also differentiated by color.

Figure 28 shows oblique views of the results draped on the DEMs. One can clearly see the crater rims (yellow) in the Columbia Hills, the dune slipfaces (cyan) in Gale Crater, and the ridges (cyan) in Mawrth Vallis. The height in these views is exaggerated by a factor of 2 for better visibility.

Figure 25. Results of cluster analysis of Columbia Hills site, 7 clusters, *t* = 5.

Table 16

Columbia Hills attribute range and mean for 7 clusters at *t* = 5

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[5,80] 32.8	[9,24] 13.7	[-0.8,0.6] -0.2	[-0.6,0.5] -0.1	hill footslope, inner crater wall
2	[34,97] 52.1	[1,20] 7.3	[-1,0.4] -0.2	[-0.9,1] 0.1	crest
3	[0,34] 10.6	[0,7] 2.2	[-0.1,0.6] 0.1	[-0.1,0.4] 0.1	floor
4	[4,83] 18.1	[1,21] 4.9	[-1,0.7] -0.1	[0.3,2.2] 0.7	crater rim
5	[0,36] 10.5	[0,8] 2.4	[-0.1,1] 0.1	[-1.2,-0.1] -0.3	floor
6	[20,82] 42.4	[5,22] 11.3	[-0.2,1.3] 0.3	[-0.9,0.5] -0.2	mid slope
7	[0,100] 49.9	[0,10] 2.2	[-1.9,-0.1] -0.3	[-0.5,0.8] 0.1	dunes
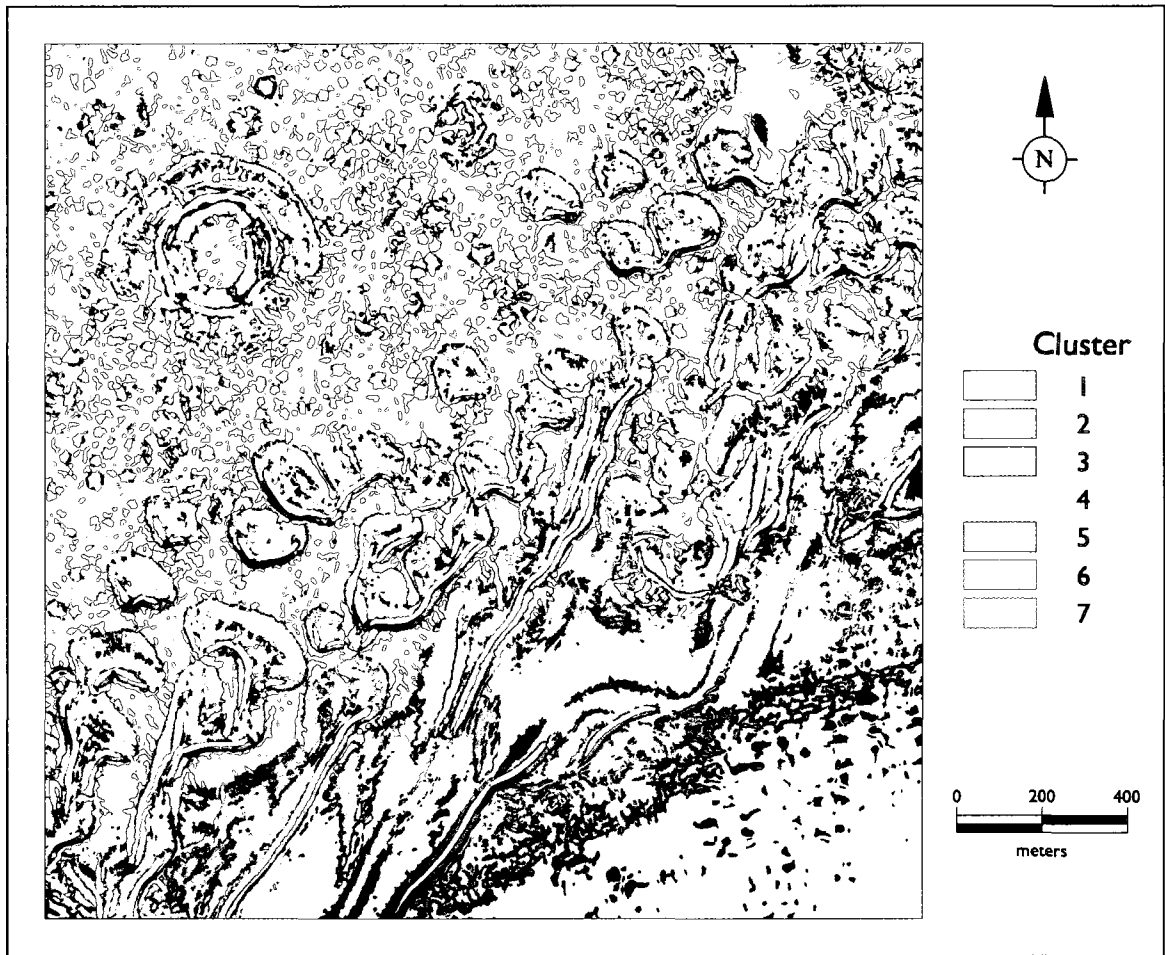
Figure 26. Results of cluster analysis of Gale Crater site at 7 clusters, *t* = 5.

Table 17

Gale Crater attribute range and mean for 7 clusters at *t* = 5

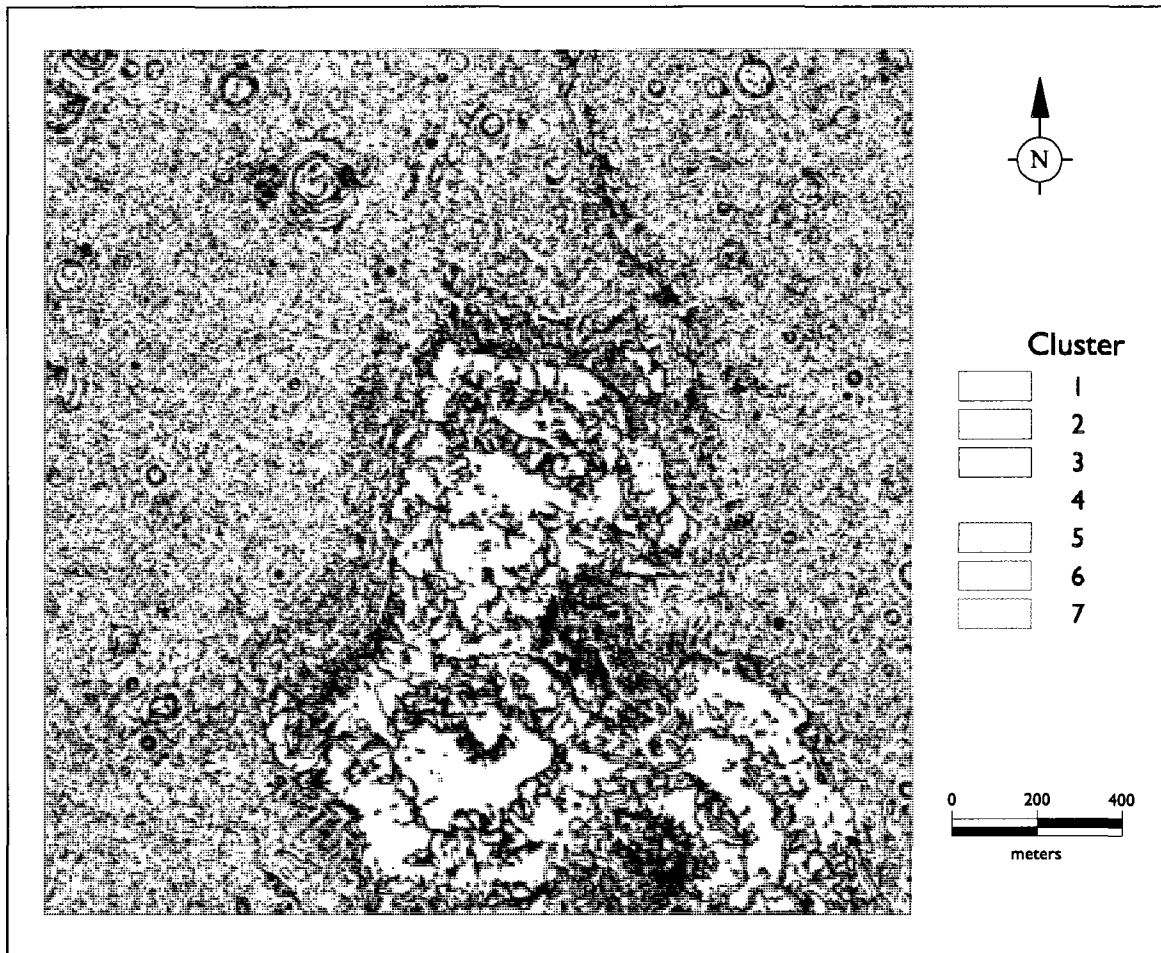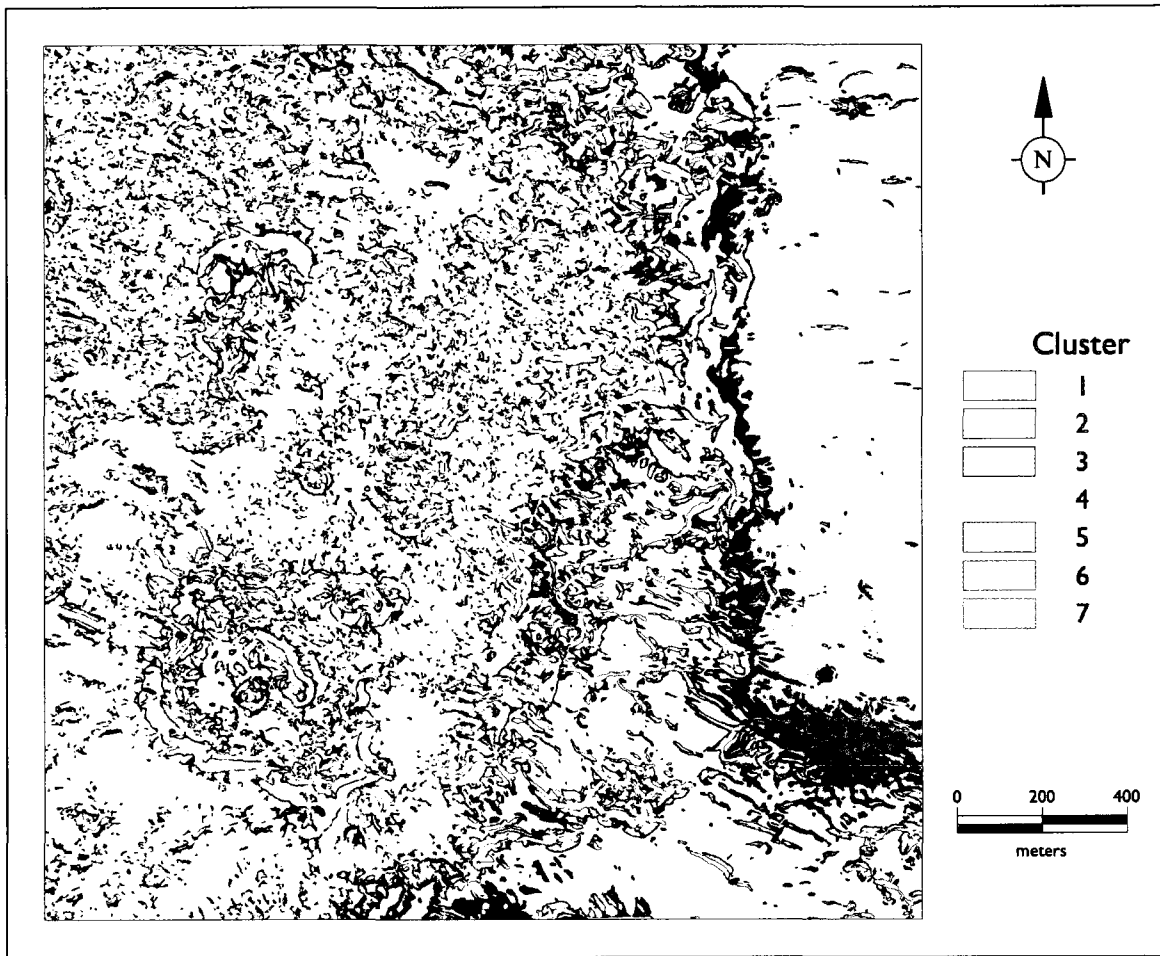| Cluster	Elevation	Gradient	Plan Curvature	Profile Curvature	Landform
1	[0,56] 12.1	[0,10] 2.6	[-1.7,-0.1] -0.4	[-0.6,1.2] 0.1	dune tops
2	[43,64] 46.6	[1.5,7] 3.8	[-0.3,0.5] 0.2	[-0.8,0.5] -0.2	upper level
3	[2,46] 14.3	[4,19] 6.7	[-0.5,0.6] 0.0	[-3,0.4] -0.3	dune, crater wall
4	[0,43] 10.2	[0,9] 2.3	[0,1.4] 0.3	[-0.9,1] -0.1	pit
5	[1,26] 9.7	[0,5] 2.4	[-0.3,0.3] 0.0	[-0.8,0.8] 0.0	floor
6	[2,58] 17	[5,18] 8.3	[-0.7,0.4] -0.1	[0.0,2.4] 0.6	rim
7	[20,51] 31.8	[1,9] 4.7	[-0.5,0.3] -0.1	[-0.8,1] 0.1	mid level

Figure 27. Results of cluster analysis of Mawrth Vallis site at 7 clusters, t = 5.

Table 18

Mawrth Vallis attribute range and mean for 7 clusters at t = 5

Cluster	Elevation	Gradient	Plan Curvature	Prof. Curvature	Landform
1	[11,93] 62.5	[9,20] 11.6	[-1,1] -0.1	[-2,0.2] -0.7	crest slope
2	[42,90] 67.7	[1,12] 5.2	[-0.2,1.4] 0.4	[-1.2,0.5] -0.2	hill lower slope
3	[15,31] 28.1	[1,13] 5.1	[-0.6,0.3] -0.1	[-0.5,0.5] 0.1	lower level
4	[31,59] 47.6	[0,10] 5.5	[-0.4,0.5] 0.0	[-1,0.6] -0.1	middle level
5	[53,91] 69.6	[1,7] 3.8	[-0.6,0.4] -0.2	[-1,0.8] -0.1	upper level
6	[4,98] 62.6	[1,25] 7.3	[-2.2,0.8] -0.3	[0.4,3.4] 1.1	rim
7	[4,96] 63	[01,25] 6.4	[-4,-0.6] -1.1	[-1.2,1.6] 0.3	convergent shoulder

55
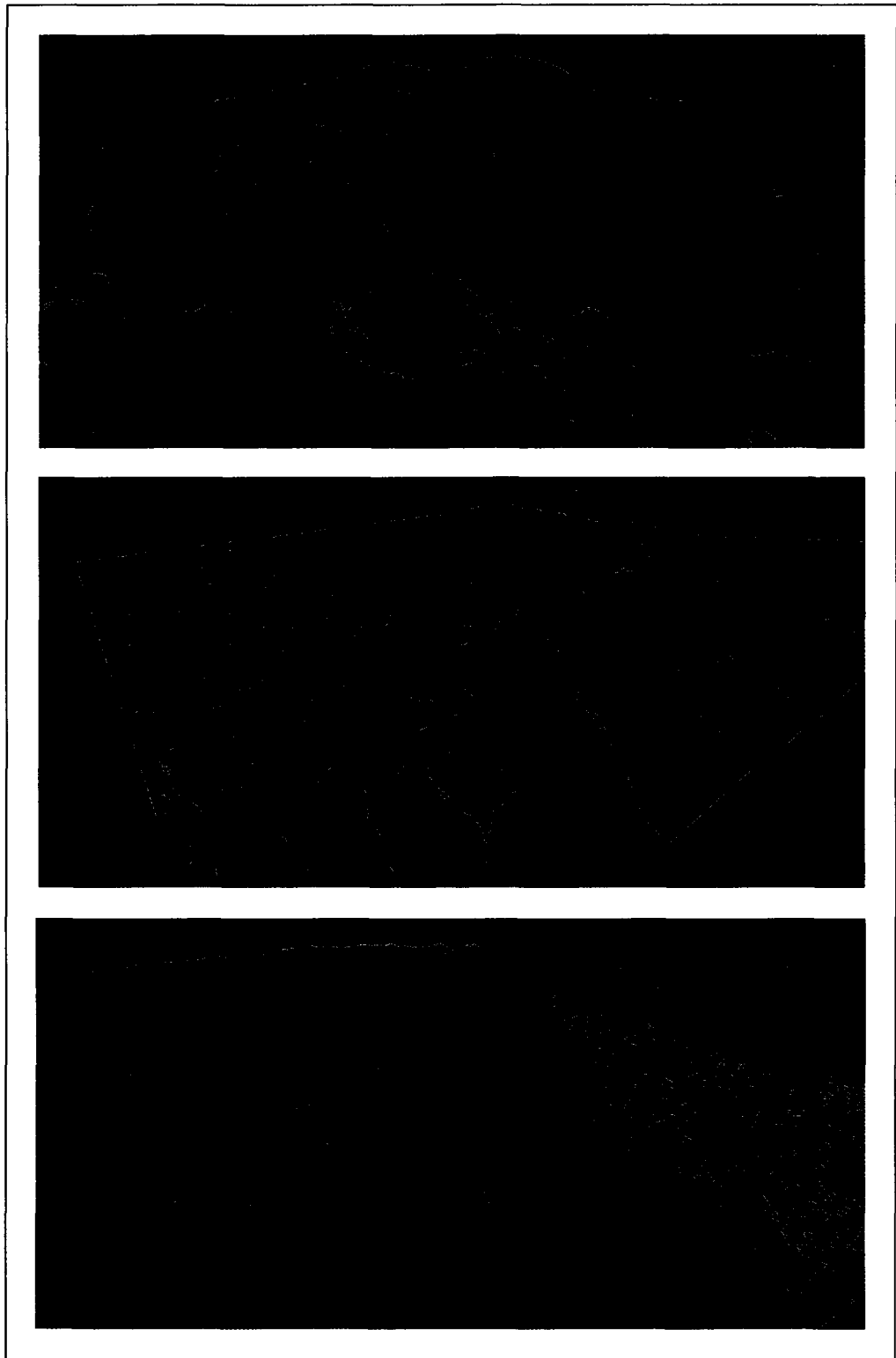
Figure 28. Oblique views of cluster analyses, 7 clusters, $t = 5$. Top: Columbia Hills, middle: Gale Crater, bottom: Mawrth Vallis (height exaggerated 2 times).

Segmentation

Figure 27 shows the Mawrth Vallis study area segmented as landform elements. This segmentation was performed on the results of a fuzzy k-mean cluster analysis with 7 clusters at scale level 5. Each element is colored separately.

One can see that there are a range of element sizes from the very large segment on the east side to the smaller elements on the peaks. The shape of the elements reflect the topography of the landscape.

Figure 29. Landform elements for Mawrth Vallis site.

Crater Detection

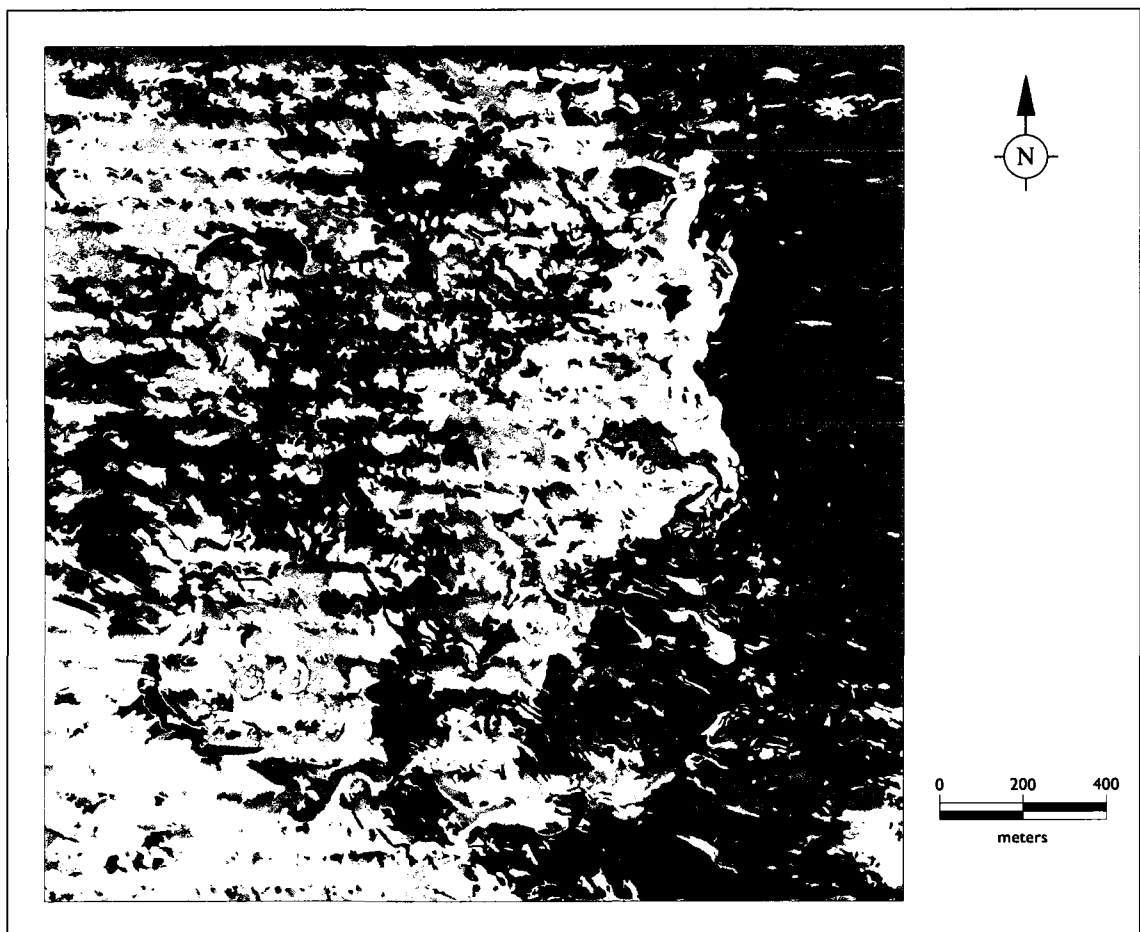Figure 30 displays the craters detected at 9 levels of scale-space $t = 2$ to $t = 10$. One can see that different craters appeared at different scale levels. Figure 31 displays all of the craters combined and overlaid on the Columbia Hills DRG.

During processing of the scale-space representations an artifact was observed. As t increased, the transition between segments become sawtoothed, which artificially increased the perimeter length of the element and affected the compactness value. A smoothing algorithm was introduced at this point to reduce this effect.

The data used in this work are very new and there is no expert crater identification available for any of these sites. As many craters as possible were visually identified from the DRG with an interactive 3D view of the DEM used as an aid. However, it was often difficult to determine what was truly a crater.

The search method found many of the craters. However, it missed those that lacked strong rims or had rims that were breached or were very close to another crater. Also, very small craters and those located at the edge of the DEM were not detected. There were also some false positives. However, upon close examination with a 3D view, it is difficult to tell if some of them may in fact be very weathered ancient craters. In this simple test, the search detected approximately 75% of the craters identified manually.
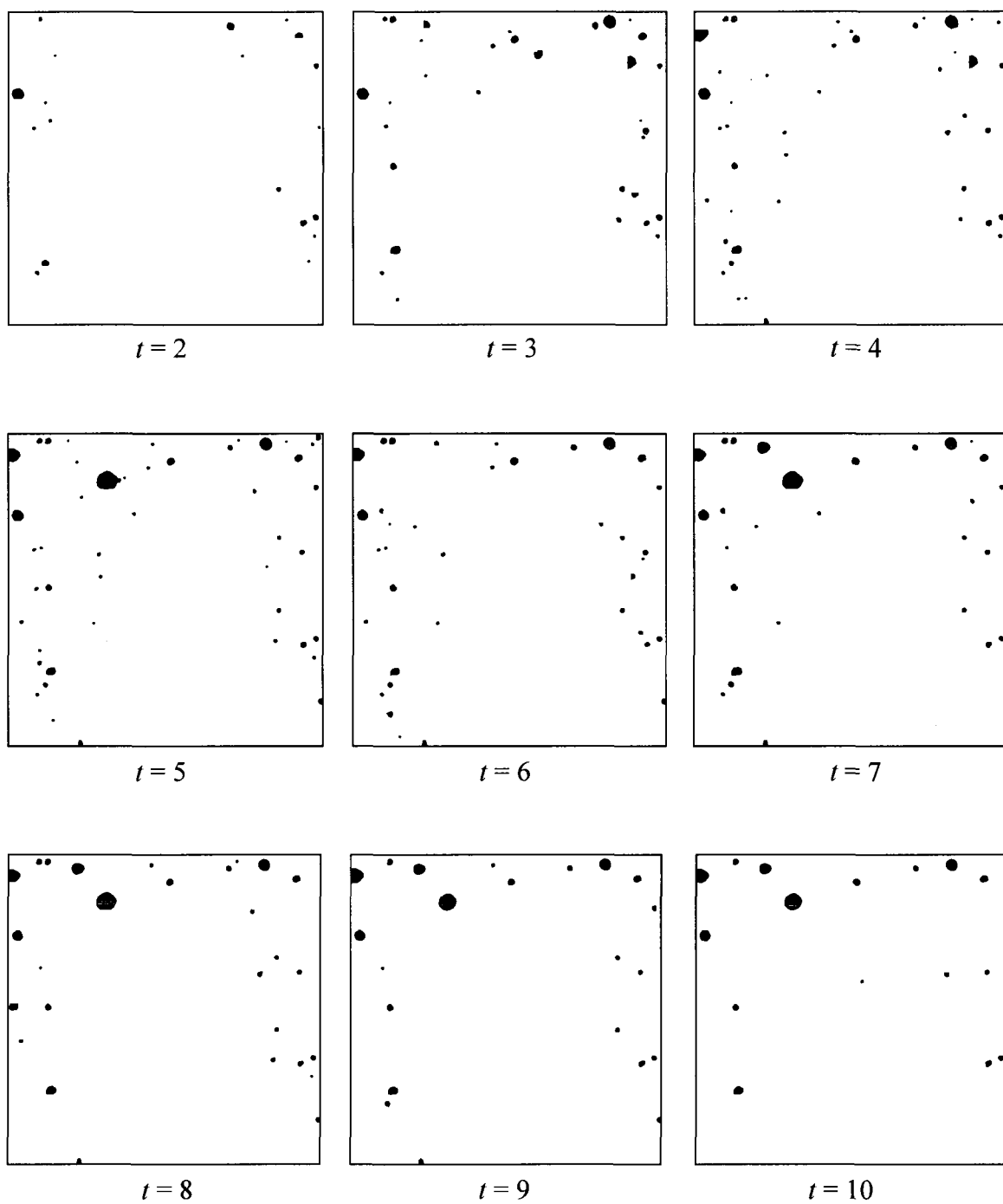
Figure 30. Craters detected at increasing scale-spaces.
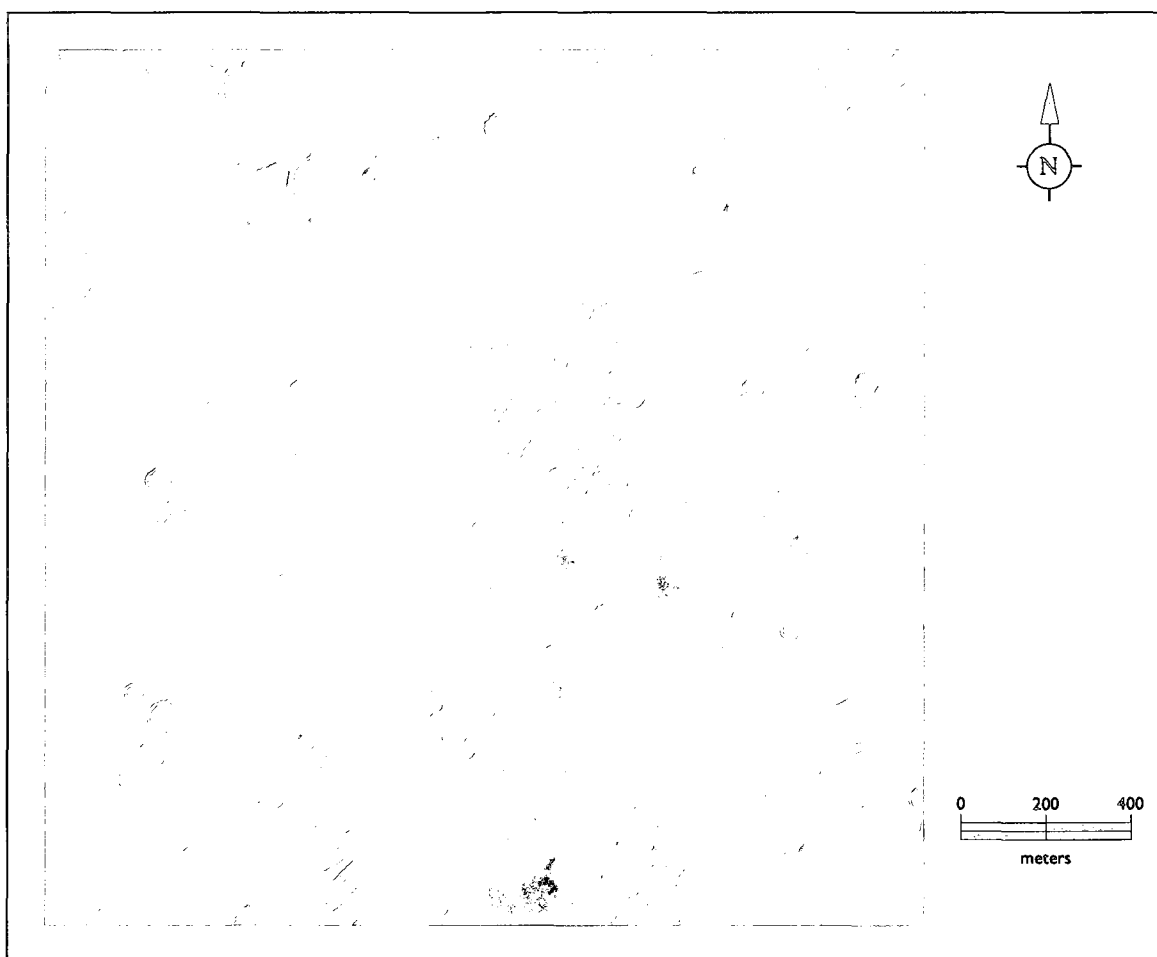
Figure 31. Craters combined and overlaid on the Columbia Hills DRG.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

Currently, Mars is accessible only through satellites, robotic explorers, and the sensors they carry. We are unable to journey to the red planet and stand on its plains or view its mountains. However, we can leverage software tools to extract every bit of usable information from these remote sources.

Mars has no vegetation and practically no atmosphere to interfere with the acquisition of HiRISE stereo images. This makes these images a prime source for building DEMs for feature extraction with fuzzy K-means. Furthermore, fuzzy K-means requires expert knowledge only in the choice of the feature vector. It does not require parameters such as hard limits on feature characteristics (e.g., size) that may affect outcome. It produces objects reflective of the most common types of features that occur in the landscape. Fuzzy K-means also allows some pixels to remain unclassified (intragrades) and it reveals the level of membership for a given cluster.

In this work, the creation of landform elements was based solely on morphology. No terrestrial based weathering or hydrological models were used. Elevation played a dominant role in two of the study sites, separating them into levels. This separation occurred in analyses with a smaller number of clusters and persisted into those with higher numbers. The high number analyses then produced clusters that were similar in every attribute but elevation, causing redundancy. Further study is necessary to

determine if this redundancy might be eliminated by replacing elevation with one or more other attributes such as contributing area or relative elevation.

A very detailed, high frequency DEM such as Columbia Hills will suffer from over-segmentation in the classification-segmentation process. Reduction of detail through the application of scale-space produced fewer, more homogeneous landform elements. Segmentation methods using homogeneity indices or a combination of homogeneity indices and scale-space may produce a better outcome.

Many improvements could be made to the crater detection technique. Using a different clustering method such as the SVM or SOM (Stepinski et al., 2006, 2007; Bue & Stepinski, 2006) or a modified feature vector with spectral components could provide a significant improvement. Additionally, results from circle identification algorithms such as the Hough transform may be helpful. An expert manual identification of craters for a HiRISE study site is very important for better validation.

The approach in this work was a general one with the objective of extracting features defined by the landscape itself. However, in the crater detection procedure and in previous work, methods focused on the characteristics of a specific type of landform to extract it from the surrounding terrain. A single general approach for all landforms types may not be possible. A combination of techniques may be necessary.

A remaining issue is the comparison of landform elements from different study sites. Allowing the landscape of a study site to define the classification raises the question of consistency between sites. Given the same feature vector and other input

parameters, will the cluster analysis for each site classify similar landform elements as the same type? The three sites in this study were quite different but the cluster analysis still identified landforms common to each site such as crater walls and hill crests. However, the defining boundary of a study site could affect the outcome of the clustering. For example, an element might be classified as a steep slope in one allocation and a shallow slope in another. The solution to this problem may require a standardized pixel classification such as that proposed by MacMillan et. al., (2000) or a standard feature vector.

In this work, landform elements from the classification-segmentation process were stored as images. Although the data are inherently raster, a feature created from these elements has a perimeter shape defined by a vector and possibly a 3D mesh. Additionally, each feature maintains object level attributes including area, width, height as well as its post level attributes such as those used for the classification. Current database and file system technologies provide venues for raster-focused or vector-focused management. Further work is required to ascertain the most optimal way to store, search, and access feature objects.

The combination of scale-space, fuzzy K-means classification, and segmentation has been shown to be a useful terrain analysis technique for Martian DEMs. It was used successfully to extract landform elements from a high detail DEM and create surface features. This simple technique is highly extensible through the modification of the

feature vector and a few parameters. With the availability of many MRO spectral data sets and their derivatives, the possibilities for future work are virtually endless.

REFERENCES

Abramoff, M. D., Magelhaes, P. J., & Ram, S. J. (2004). Image Processing with ImageJ. *Biophotonics International,* 11, 36-42.

Arrell, K. E., Fisher, P. F., Tate, N. J., & Bastin, L. (2007). A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in Snowdonia, Wales. *Computers & Geosciences,* 33, 1366-1381.

Baatz, M. & Schape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In J. Strobl, T. Blaschke, & G. Griesebner (Eds.) *Angewandte Geographische Informationsverarbeitung XII* (pp. 12-23). Heidelberg: Wichmann-Verlag.

Bell, J., Edgett, K., Rowland, S. & Malin, M. (2006). *The Gale Crater Mound: A Candidate Landing Site for the 2009 Mars Science Laboratory.* Retrieved March 21, 2009 from http://marsoweb.nas.nasa.gov/landingsites/msl/workshops/ 1st_workshop/program.html

Bezdek, J. C., Ehrlich, R. & Full, W. (1984). FCM: The Fuzzy c-Means Clustering Algorithm. *Computers & Geosciences,* 10(2,3), 191-203.

Bue, B. D. & Stepinski, T. F. (2006). Automated classification of landforms on Mars. *Computers & Geoscience,* 32(5), 604-614.

Bue, B.D. & Stepinski, T. F. (2007). Machine Detection of Martian Impact Craters From Digital Topography Data. *IEEE Transactions on Geoscience and Remote Sensing,* 45(1), 265-274.

Burrough, P. A., van Gaans, P.F. M. & MacMillan, R. A. (2000). High-resolution landform classification using fuzzy K-means. *Fuzzy Sets and Systems,* 113, 37-52.

de Gruijter, J. J. & McBratney, A. B. (1988). A Modified Fuzzy K-means Method for Predictive Classification. In H. H. Bock (Ed.) *Classification and Related Methods of Data Analysis* (pp. 97-104). Amsterdam: Elsevier Science.

de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2007). *Geospatial Analysis A Comprehensive Guide to Principles, Techniques and Software Tools*. London: Winchelsea Press.

Dragut, L. & Blaschke, T. (2006). Automated classification of landform elements using object-based image analysis. *Geomorphology, 81*, 330-344.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: Wiley & Sons.

Easterbrook, D. J. (1999). *Surface Processes and Landforms* (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Grant, J. A. et. al. (2004). Selecting landing sites for the 2003 Mars Exploration Rovers. *Planetary and Space Science, 52*, 11-21.

Jensen, J. R. (2005). *Introductory Digital Image Processing A Remote Sensing Perspective* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall

Jet Propulsion Lab. (2008, April 30). *Mars Reconnaissance Orbiter*. Retrieved March 21, 2009 from http://mars.jpl.nasa.gov/mro

Jet Propulsion Lab. (2009, March 20). *Mars Exploration Rover Mission*. Retrieved March 21, 2009 from http://marsrovers.nasa.gov/home/index.html

Li, Z., Zhu, Q., & Gold, C. (2005). *Digital Terrain Modeling Principles and Methodology*. Boca Raton, Florida: CRC Press.

Lindeberg, T. (1996). Scale-space: A framework for handling image structures at multiple scales. *Proceedings of the Cern School of Computing*, The Netherlands, September, 1996.

MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., & Goddard, T. W. (2000). A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy Sets and Systems, 113*, 81-109.

Mangan, A. P. & Whitaker, R. T. (1999). Partitioning 3D Surface Meshes Using Watershed Segmentation. *IEEE Transactions on Visualization and Computer Graphics, 5*(4), 308-321.

Michalski, J. R. et. al. (2006). *Clay Minerals in the Mawrth Vallis Region: A proposed MSL landing site*. Retrieved March 21, 2009 from http://marsoweb.nas.nasa.gov/landingsites/msl/workshops/2nd_workshop/program.html

Molloy, I. & Stepinski, T. F. (2007). Automatic Mapping of Valley Networks on Mars. *Computers & Geosciences*, 33(6), 728-738.

NASA Ames Research Center. (2008). *2009 Mars Science Laboratory*. Retrieved March 21, 2009 from http://marsoweb.nas.nasa.gov/landingsites/index.html

Pyle, D. (1999). *Data Preparation For Data Mining*. San Francisco, CA: Morgan Kaufmann.

Rogerson, P. A. (2006). *Statistical Methods for Geography A Student's Guide* (2nd ed.). London: Sage.

Squyres, S. (2000). *Roving Mars*. New York: Hyperion.

Stepinski, T. F. & Vilalta, R. (2005). Digital Topography Models for Martian Surfaces. *IEEE Geoscience & Remote Sensing Letters*, 2(3), 260-264.

Stepinski, T. F., Ghosh, S., & Vilalta, R. (2006). Automatic Recognition of Landforms on Mars Using Terrain Segmentation and Classification. In Lecture Notes in Computer Science: Vol. 4265. *Discovery Science* (pp. 255-266). Heidelberg: Springer Berlin.

Stepinski, T. F., Ghosh, S. & Vilalta, R. (2007). Machine Learning for Automatic Mapping of Planetary Surfaces. In Proceedings of 19th Innovative Applications of Artificial Intelligence Conference (pp. 7-18). AAAI Press, 2007.

USGS. (2009). *Planetary GIS Web Server*. Retrieved November 17, 2008 from http://webgis.wr.usgs.gov

University of Arizona. (2009). *High Resolution Imaging Science Experiment*. Retrieved April 18, 2009 from http://hirise.lpl.arizona.edu

Ventura, S. J., & Irvin, B. J. (2000). Automated Landform Classification Methods for Soil-Landscape Studies. In J. P. Wilson & J. C. Gallant (Eds.), *Terrain Analysis Principles and Applications* (pp. 267-294). New York: Wiley & Sons.

Wilson, J. P. & Gallant, J. C. (Eds.). (2000). *Terrain Analysis Principles and Applications*. New York: Wiley & Sons.

Zevenbergen, L. W. & Thorne, C. R. (1987). Quantitative Analysis of Land Surface Topography. *Earth Surface Processes and Landforms,* 12, 47-56.