CS 739 Spatial Data Mining

Term Project: Crater Classification on Mars Crater Data Set

Assigned Date: Sunday, March 14, 2011

Educational Goal

Apply Artificial Intelligence and Machine Learning techniques to real-world Mars crater data.

Team

Henry Z. Lo, Shun Guang, Joseph Paul Cohen

Phase III: Training Set Construction and Data Visualization

(200 points)

Assigned Date: Tuesday, April 5, 2011

Due: 5:30 PM Tuesday, April 19, 2011

Requirements

- **Sampling of Negative Examples:** Explore different ways of generating negative examples. Include images that contain partial craters. Find out of the classifier needs more examples of a certain percentage of crater overlap than others. Empirically test each training set generated, and find the one with the best results.
- **ArcGIS Visualization:** Create 4 visualization layers in ArcGIS: the image layer, the true positive layer, the false positive layer, and the false negative layer. Create points on the result layers to visualize craters, and make the size of the points relative to the size of the crater detected. Make sure the points are transparent, so that the craters themselves are still visible.
- **Generate Results and Analyze:** In order to analyze results, output the coordinates and the size of each crater detected. Also, find a method of matching detected craters with the ground truth, in order to determine whether it is a FP, TP, or FN.
- **Understand the OpenCV Code:** In the next phase, we may want to use the same Haar features as OpenCV in order to compare algorithms effectively. To do this, we need to understand how OpenCV creates and uses Haar features in its cascade classification algorithm and export these features for usage elsewhere.
- **Find Ideal Range for Crater Detection:** Plot accuracy as a function of crater size. Determine the effective range of our crater classifier. The upper size limit of craters to identify is 4 km.

CS 739 Spatial Data Mining

Submission Requirements

1. Generate a report of the results of every test run. In this report, include side-by-side comparisons of techniques and a side-by-side comparison of data sets.

2. Write a brief summary on the effectiveness of preprocessing, differences in classification algorithms, and the relative utility of each DEM and optical data. Also identify future directions of research.

3. Submit reports through Blackboard.