

## Term Project: Detection of Small Craters in High Resolution Planetary Images Using Shape and Texture Features

### Educational Goal

Apply spatial data mining techniques to real-world applications.

### Phase I-Supplementary Assignment : Training Set Analysis (100 points)

**Due: 8:30 PM Thursday April 15, 2010**

### Requirements

- Use the same training set in Phase I.
- 5 test sets are provided for this assignment. Download the files from Term Project Phase I: information package via UMass Online. The 5 test sets are tiles 1\_24, 1\_25, 2\_24, 2\_25, and 3\_25. The training set is sampled from tile 3\_24 and tile 3\_24 is NOT used for testing in this assignment.
- The arrangement of the 6 tiles in **Error! Reference source not found.** is:

Tile 1_24	Tile 2_24	Tile 3_24
Tile 1_25	Tile 2_25	Tile 3_25



Figure 1 Site for the study, located in the Xanthe Terra, centered on Nanedi Vallis and covers mostly Noachian terrain. The site's north-south extension is 40 km, east-west extension is nearly 60 km and it was divided into six overlapping tiles.

- Test set format. The 1st and 2nd columns are the x and y coordinates of the centroid of a crater candidate (measurement unit is pixel); the 3rd column is the radius of the crater candidate (measurement unit is pixel); columns 4 – 1092 are the 1089 texture-based features of the crater candidate (data type is double); column 1093 is the class label, where 1 is for crater and 0 is for non-crater.
- Use Weka J48 to build a decision tree from the training set and perform classification on the 5 test sets. For each test set, design 20 experiments to gradually build Underfitting decision trees to overfitting decision trees. Report accuracy, error rate, precision, recall, and F-measure for the 20 experiments. There are two ways to control pruning in decision trees using Weka J48: confidence level and the reducedErrorPruning scheme. Please refer the online article for details on those two schemes: <http://condor.depaul.edu/~ntomuro/courses/578/assign/hw1.html>
- Implement a script to call Weka J48 API to run the 20 experiments and **automatically** extract the experiments results on accuracy, error rate, precision, recall, and F-measure. 0 points if the results are extracted manually.

### Submission Requirements

1. Write an experiment report to discuss all the experiments. In the report, discuss the design of experiments and experiment results in 3 regions, and your conclusion. The 5 tiles have different geographical features. The east region tile 3\_25, the central region tiles 2\_24 and 2\_25, and the west region tiles 1\_24 and 1\_25 may exhibit different characteristics with respect to training errors and test errors. Evaluate your experimental results carefully and discuss any interesting results.
2. Submit the paper copy of the report, and source code of the scripts with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.
3. Submit the softcopy of the report and scripts through your UMassOnline account at <http://boston.umassonline.net/index.cfm>.
4. Zip all the files. One submission per team. Save the file as sdm\_teamNumber. For example, Team 1 should name their file as *sdm\_team1.zip*.
5. No hard copies or soft copies results in 0 points.