

An Effective Approach for Selection of Terrain Modeling Methods

Guimin Jia, Xiangjun Wang, and Hong Wei

Abstract—This letter presents an effective approach for selection of appropriate terrain modeling methods in forming a digital elevation model (DEM). This approach achieves a balance between modeling accuracy and modeling speed. A terrain complexity index is defined to represent a terrain's complexity. A support vector machine (SVM) classifies terrain surfaces into either complex or moderate based on this index associated with the terrain elevation range. The classification result recommends a terrain modeling method for a given data set in accordance with its required modeling accuracy. Sample terrain data from the lunar surface are used in constructing an experimental data set. The results have shown that the terrain complexity index properly reflects the terrain complexity, and the SVM classifier derived from both the terrain complexity index and the terrain elevation range is more effective and generic than that designed from either the terrain complexity index or the terrain elevation range only. The statistical results have shown that the average classification accuracy of SVMs is about $84.3\% \pm 0.9\%$ for terrain types (complex or moderate). For various ratios of complex and moderate terrain types in a selected data set, the DEM modeling speed increases up to 19.5% with given DEM accuracy.

Index Terms—Support vector machine (SVM), terrain classification, terrain complexity, terrain modeling.

I. INTRODUCTION

A DIGITAL elevation model (DEM) is a digital representation of terrain topography. Original elevation data can be obtained from either ground survey or passive and active remote sensing [1]–[3] in a discrete form. To reconstruct a continuous terrain surface from these discrete elevation data, two types of data structures, i.e., “grid” and “triangular irregular network (TIN),” are popularly used. The grid data structure has many advantages such as simple for storage, and relevant terrain modeling algorithms tend to be simple and straightforward. However, the uniform data density has limits in coping with the complexity of land surface. The TIN data structure has changeable resolutions that adapt to different terrain types and may need less storage space. The adaptability of data density

Manuscript received June 28, 2012; revised October 9, 2012; accepted October 17, 2012. Date of current version February 20, 2013. This work was supported in part by the National High Technology Research and Development Program 863 under Grant 2010AA122206-4, by the Key Technologies R&D Program of Tianjin, China, under Grant 08ZCKFJC27900, and by the Natural Science Foundation of China under Grant 60872097.

G. Jia and X. Wang are with the State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China, and also with the Micro Optical Electro Mechanical Systems Education Ministry Key Laboratory, Tianjin University, Tianjin 300072, China (e-mail: gm_j1982@yahoo.com.cn).

H. Wei is with the School of Systems Engineering, University of Reading, Berkshire RG6 6AY, U.K.

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2012.2226429

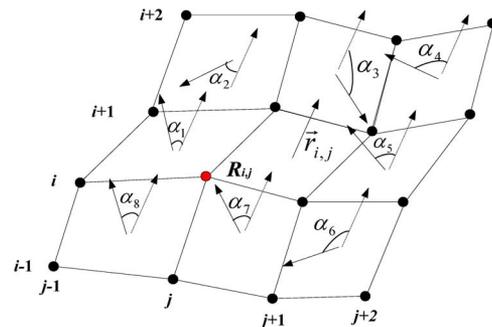


Fig. 1. Sketch map of a terrain complexity index.

makes a TIN more accurate in modeling a terrain surface. However, the TIN data structure usually takes longer time in the reconstruction of DEM and is difficult for surface analysis [4], [5]. In practice, it always needs to compromise between DEM accuracy and time used for modeling. The choice of terrain modeling algorithms mainly depends on data availability, land surface characteristics, and the required DEM accuracy. Once the source data are determined, the nature of land surface will have significant influence on the selection of terrain modeling algorithms and further affect the DEM accuracy and modeling speed.

This letter presents an effective approach for selection of terrain modeling algorithms based on the natural complexity of land surface. Between “grid” and “TIN,” the basic rule is that a grid-based algorithm is selected by default, and the TIN-based algorithm is only used when the grid-based algorithm does not satisfy the required DEM accuracy. In this letter, a terrain complexity index is defined to represent terrain characteristics, the terrain modeling error is used to measure DEM accuracy, and a support vector machine (SVM) is applied to classify terrains into two classes, namely, complex (unsuitable for grid) and moderate (suitable for grid). Finally, the appropriate terrain modeling method is recommended, which balances the modeling accuracy and the modeling speed. Sample terrain data from the lunar surface are used in the experiment, and the results prove that the proposed approach is practicable for selection of terrain modeling algorithms.

II. QUANTITATIVE ASSESSMENT OF TERRAIN COMPLEXITY

There is no uniform and optimal measure for quantitatively assessing terrain complexity [6], [7]. We define a terrain complexity index for this purpose. As shown in Fig. 1, taking a 3×3 space analysis window, the average angle among the center grid and its eight neighbor grids is related to its terrain complexity.

In Fig. 1, $R_{i,j}$ represents the elevation point in row i and column j and the grid taking this point at the bottom left corner, $\vec{r}_{i,j}$ is the unit normal vector of grid $R_{i,j}$, and $\vec{r}_{i,j} = (a_{i,j}, b_{i,j}, c_{i,j})$; $\alpha_1, \alpha_2, \dots, \alpha_8$ is the angle between the center grid and the eight neighbor grids.

In Fig. 1, we can see that with the increase in α values, the terrain change between the adjacent grids increases, too. α is defined in the interval of $[0^\circ, 180^\circ]$.

A. Definition of the Terrain Complexity Index

In order to establish the terrain complexity index, a conversion factor is introduced as follows:

$$Tc_\alpha = e^{-\cos \alpha}. \quad (1)$$

Taking $R_{i,j}$ and $R_{i,j-1}$, for example, Tc_α is calculated according to (2) and (3), i.e.,

$$\cos \alpha_1 = a_{i,j}a_{i,j-1} + b_{i,j}b_{i,j-1} + c_{i,j}c_{i,j-1} \quad (2)$$

$$Tc_{\alpha_1} = e^{-(a_{i,j}a_{i,j-1} + b_{i,j}b_{i,j-1} + c_{i,j}c_{i,j-1})}. \quad (3)$$

Equations (2) and (3) indicate that Tc_α increases with the increase in α in the range of $[0^\circ, 180^\circ]$. A large Tc_α indicates a large change between adjacent grids. Consequently, the terrain complexity index for grid $R_{i,j}$ can be calculated as follows:

$$TCI_{i,j} = \frac{Tc_{\alpha_1} + Tc_{\alpha_2} + \dots + Tc_{\alpha_8}}{8} \quad (4)$$

For an area that consists of $M \times N$ grids, each $TCI_{i,j}$ ($i = 1, 2, \dots, M; j = 1, 2, \dots, N$) expresses a local terrain change, and for the whole area under consideration, the terrain complexity index is defined as

$$E_{TCI} = \frac{\sum_{i=1}^M \sum_{j=1}^N TCI_{i,j}}{MN}. \quad (5)$$

With the increase in E_{TCI} , the terrain surface variation increases, too. In this letter, the terrain complexity is expressed by E_{TCI} . On the basis that the required DEM accuracy is satisfactory, the space analysis window could be set to 5×5 , 7×7 , or even larger, when the sampling interval of elevation points is very small. This can speed up the E_{TCI} calculation. Since the DEM resolution may affect E_{TCI} , it is necessary to use the same DEM resolution in the comparison of the terrain complexity by E_{TCI} .

B. Proof of the Correctness of E_{TCI}

Experiments are conducted to prove the correctness of the defined E_{TCI} . A function popularly used for simulating terrain surfaces [8] is adopted, as shown in (6). By adjusting the coefficients A , B , and C in (6), terrain surfaces with different complexities can be generated, i.e.,

$$Z = A \times \left[1 - \left(\frac{X}{m} \right)^2 \right] \times e^{-\left(\frac{X}{m} \right)^2 - \left(\frac{Y}{n} + 1 \right)^2} - B$$

TABLE I
COMPARISON OF TERRAIN COMPLEXITY INDEXES OF
TERRAINS WITH DIFFERENT COMPLEXITIES

Coefficient value			Elevation Range	E_{TCI}
A	B	C	(unit: m)	
10	20	0.5	-10~16	0.3496
50	20	0.5	-10~40	0.3497
50	120	0.5	-80~100	0.3499
50	120	3	-100~75	0.3499
100	240	8	-125~200	0.3503
200	360	50	-200~300	0.3511
500	400	120	-300~400	0.3520
700	1100	340	-800~900	0.3562

$$\times \left[0.2 \times \frac{X}{m} - \left(\frac{X}{m} \right)^3 - \left(\frac{Y}{n} \right)^5 \right] \times e^{-\left(\frac{X}{m} \right)^2 - \left(\frac{Y}{n} \right)^2} - C \times e^{-\left(\frac{X}{m} + 1 \right)^2 - \left(\frac{Y}{n} \right)^2} \quad (6)$$

where A , B , and C are topographic relief parameters; m and n are range control parameters; and X , Y , and Z are coordinates of a ground point.

The range of the simulated terrain is $\{X \in [-1000 \text{ m}, 1000 \text{ m}], Y \in [-1000 \text{ m}, 1000 \text{ m}]\}$, and the spatial resolution of the DEM model is 50 m. By changing the values of A , B , and C gradually, eight simulated terrain surfaces are generated with different complexities. According to (3), their terrain complexity index is calculated in Table I.

Table I indicates that the terrain complexity is positively proportional to the terrain complexity index.

Fig. 2 shows two DEMs (the spatial resolution is 30 m) of the lunar surface, i.e., LOLA_038 and LOLA_043 (the data details are introduced in Section V-A). The elevation ranges of LOLA_038 and LOLA_043 are about 3000 and 220 m, respectively; and LOLA_043 is obviously more complex than LOLA_038. According to (1)–(5), the terrain complexity indexes of LOLA_038 and LOLA_043 are 0.3704 and 0.3865, respectively. Therefore, we have the conclusion that LOLA_043 is a more complex terrain than LOLA_038. This further proves that the terrain complexity index represents the terrain characteristic in terms of complexity.

III. TERRAIN REPRESENTATION ERROR

In this letter, we use terrain representation error to represent the difference between a DEM and a real elevation [9]. On the assumption that the sampling error of elevation points is zero, the difference between the terrain model and the actual terrain is defined as the terrain representation error. For the elevation point $R_{i,j}$, the terrain representation error can be calculated as follows:

$$E_{t(i,j)} = H_{(i,j)} - \left(H_{(i-1,j-1)} + H_{(i-1,j+1)} + H_{(i+1,j-1)} + H_{(i+1,j+1)} \right) / 4 \quad (7)$$

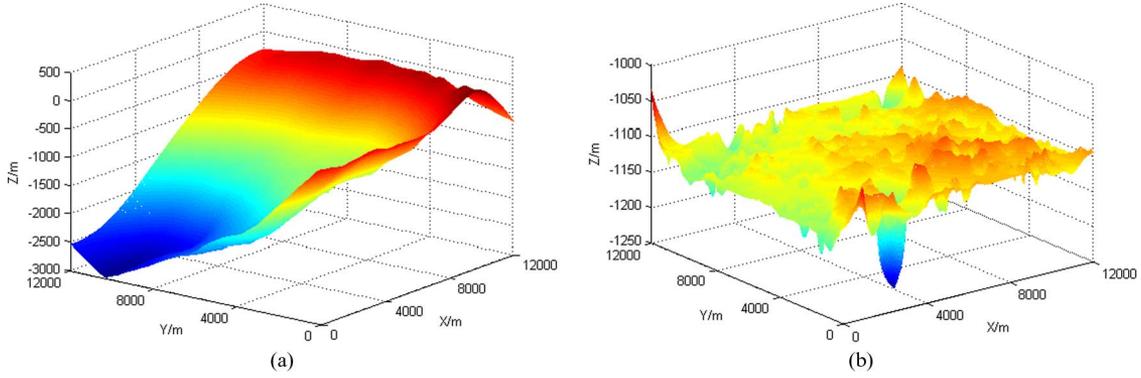


Fig. 2. Comparison of lunar terrains with different terrain complexities. (a) LOLA_038 ($E_{TCI} = 0.3704$). (b) LOLA_043 ($E_{TCI} = 0.3865$).

where $H_{(i,j)}$ is the real elevation at point $R_{i,j}$ (true data), and the second part in (7) gives the calculated elevation of point $R_{i,j}$ based on the DEM. For an area that consists of $M \times N$ grids (note: in this case, the DEM has been downsampled to $M/2 \times N/2$), the terrain representation error can be calculated by the root-mean-square error of E_t as follows:

$$\text{RMSE}_{E_t} = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (E_{t(i,j)} - E(E_t))^2}{MN - 1}} \quad (8)$$

where $E(E_t)$ represents the mean value of E_t across a considered range.

IV. TERRAIN CLASSIFICATION BASED ON AN SVM

Equations (1)–(5) calculate the terrain complexity index, and then, an SVM classifies terrains into either complex or moderate in association with the terrain elevation range. The approach for selection of DEM modeling algorithms is formed based on the classification result. In this letter, the data source used in experiments is regularly sampled data (we will introduce the data source in Section V-A), so that the default DEM is “grid” and that “TIN” stands as the optional. The detailed algorithm for terrain classification is described as follows:

- 1) choose N different terrain surfaces, and calculate their terrain complexity indexes: E_{TCI_i} , $i = 1, 2, \dots, N$;
- 2) model the chosen terrains using the grid-based algorithm, and calculate the terrain representation error for each terrain surface according to (8): $\text{RMSE}_{E_{ti}}$, $i = 1, 2, \dots, N$;
- 3) form a sample pool from the N terrain surfaces, set a threshold T_h for $\text{RMSE}_{E_{ti}}$:

$$\text{Positive samples : } \text{RMSE}_{E_{ti}} > T_h$$

$$\text{Negative samples : } \text{RMSE}_{E_{ti}} \leq T_h$$

- 4) use TM_i , $i = 1, 2, \dots, N$ to express the classification result:

$$TM_i = \begin{cases} 1, \text{ positive sample, complex terrain,} \\ \quad \text{TIN is recommended} \\ 0, \text{ negative sample, moderate terrain,} \\ \quad \text{grid applicable} \end{cases}$$

- 5) take E_{TCI} (with or without elevation range) and TM_i as the input and output of the SVM, respectively; and select

samples randomly and successively from the positive and negative samples to compile the training set and the testing set;

- 6) train the SVM using the training set, and evaluate the SVM classifier by the testing set.

In choosing the threshold T_h , the following equation is used:

$$T_h = \frac{\text{RMSE}_{E_{t1}} + \text{RMSE}_{E_{t2}} + \dots + \text{RMSE}_{E_{tN}}}{N} \cdot k \quad (9)$$

where $k \in [0, 1]$ is a relaxation coefficient, which is set according to the required DEM accuracy. When high accuracy is required, k should be set to a small value; otherwise, k could be a larger value.

V. DATA SOURCE AND SVM CLASSIFICATION EXPERIMENTS

A. Data Source

Data sources for DEM generation, in general, are from the following three classes: those collected from 1) ground survey; 2) existing topographic maps; and 3) remote sensing. Our research project requires the DEM reconstruction for a large area of a lunar terrain. In consideration of both the data availability and the data accuracy, the Lunar Orbiter Laser Altimeter (LOLA) data, which were released by National Aeronautics and Space Administration, are taken as the source data in testing experiments [10]. They are regular grids with a resolution of 30 m.

B. Experiments of SVM Terrain Classification

Eighty tiles of lunar terrain surfaces with different features are selected, and the range covers 12000 m \times 12000 m for each tile. Fig. 3 shows the study area in this letter. The terrain tiles are randomly selected from the study area. In Fig. 3, A, B, C, and D are four examples among the 80 terrain tiles.

An SVM classifier is trained by using LIBSVM tool box, which was mainly developed by C.J Lin from National Taiwan University. LIBSVM is integrated software for SVM classification, regression, and distribution estimation. In this letter, version 2.83 of LIBSVM was used, and experiments were carried out in MATLAB. In training the SVM classifier, a radial basis kernel function is adopted.

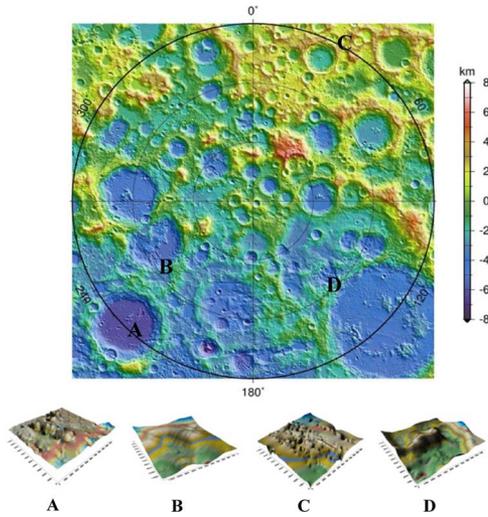


Fig. 3. Study area—lunar south pole elevations: -75° to the South Pole, by the Lunar Reconnaissance Orbiter LOLA Science Team, 30 m/pixel.

It has been noticed that the terrain complexity index of some terrains such as LOLA_038 is relatively high due to the large elevation variation of the data, but it does not naturally refer to a complex terrain. Both the complexity index and the elevation range are involved in the terrain type classification. Three SVMs are established—they are as follows: 1) SVM_ E_{TCI} , which is the SVM based only on the E_{TCI} feature; 2) SVM_Elevation, which is the SVM based only on the feature of the elevation range; and 3) SVM_ E_{TCI} + Elevation, which is the SVM based on both features.

The 80 samples construct the positive and negative data sets. The positive data set with 27 samples refers to complex terrain surfaces for which a grid-based modeling method cannot meet the required DEM accuracy, whereas the negative data set contains 53 samples in which terrain surfaces are moderate. The training set consists of 20 positive samples and 40 negative samples, and the testing set consists of 7 positive samples and 13 negative samples. Values (the terrain complexity index and the terrain elevation range) in the input matrix are normalized for training and testing in SVMs.

1) *Classification Accuracy*: The terrain type classification accuracy by an SVM is defined in (10), shown at the bottom of the page. Random combinations of the training sets (60 training samples from the 80 tiles), i.e., 50 test times, give 50 classification results for the three models (i.e., SVM_ E_{TCI} , SVM_Elevation, and SVM_ E_{TCI} + Elevation), which are shown in Fig. 4. The statistic result from the 50 tests is presented in Table II. *MSE* in Table II is defined by (11), shown at the bottom of the page, which gives the stability of the classification accuracy of SVMs. From Fig. 4 and Table II,

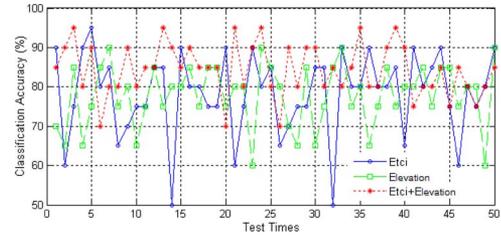


Fig. 4. Accuracy comparison of the three SVMs at 50 tests.

TABLE II
PREDICTION RESULTS OF THE THREE SVMs FOR THE TESTING SET

SVMs	Average classification accuracy (%)	MSE
SVM_ E_{TCI}	78.4	1.4727
SVM_Elevation	77.9	1.1233
SVM_ E_{TCI} +Elevation	84.3	0.9382

TABLE III
CONFUSION MATRIX OF TEST 3

True class	SVM_ E_{TCI} +Elevation	
	Predicted true	Predicted false
True	7	0
False	1	12

we can conclude that, by combining E_{TCI} and the elevation range as the input features to the SVM model, the average classification accuracy achieves $84.3\% \pm 0.9\%$ for terrain types, namely, complex or moderate. The results demonstrate that the classification accuracy of SVM_ E_{TCI} + Elevation is higher and more stable than the other two models. Confusion matrices of tests 3, 20, and 47 of SVM_ E_{TCI} + Elevation are shown in Tables III–V, respectively. By using this method, the classification error is mainly caused by training samples when the structure of SVMs is fixed.

DEM Modeling Speed: Computing specifications in the experiments are as follows:

- Intel Pentium dual core G630 2.70 GHz, 4-G DDR3 memory, 500-G SATA hard disk, and 512-M independent ATI Radeon graphics card;
- The software used in the DEM modeling is AutoCAD 2006, and SVM classification is carried out with LIBSVM 2.83 in MATLAB R2010a.

$$\text{classification accuracy} = \frac{\text{number of samples correctly classified}}{\text{total number of samples}} \tag{10}$$

$$MSE = \frac{\sqrt{\sum_{n=1}^{50} (\text{classification accuracy} - \text{average classification accuracy})^2}}{50} \tag{11}$$

TABLE IV
CONFUSION MATRIX OF TEST 20

True class	SVM_ E_{TCI} + Elevation	
	Predicted true	Predicted false
True	5	2
False	4	9

TABLE V
CONFUSION MATRIX OF TEST 47

True class	SVM_ E_{TCI} + Elevation	
	Predicted true	Predicted false
True	6	1
False	3	10

In order to show the modeling speed gained from our method, a comparison experiment is conducted with 20 samples in the testing set, as shown here.

- 1) The DEM modeling without using our method (the overall time spent is marked as T_{direct}):
 - modeling 20 terrain tiles by the grid-based modeling method;
 - calculating $RMSE_{E_{ti}}$ ($i = 1, 2, \dots, 20$) for each DEM model, if $RMSE_{E_{ti}} > T_h$, the TIN-based method is used to model the terrain.
- 2) The DEM modeling by using our method (the overall time spent is marked as T_{with_SVM}):
 - calculating E_{TCH} and the elevation range for each terrain tile;
 - classifying the terrains by SVMs;
 - modeling these terrains by the grid-based or TIN method according to the classification result;
 - to avoid the effect of the SVM classification error on the DEM accuracy, calculating $RMSE_{E_t}$ for each DEM model that is established by the grid-based modeling method, if $RMSE_{E_t} > T_h$, the TIN-based method is reapplied to model the terrain.

It is obvious that when the ratio of complex terrain in a data set is high, the proposed approach is more effective. However, how effective will it be? With different ratios of complex and moderate terrains in a data set, four combinations are used in the experiment to answer the question. The results are presented in Table VI.

The DEM modeling speed of our method is faster in comparison with the direct modeling method. With the increase in the number of positive samples in the data set, our proposed strategy shows a higher speed gain in DEM modeling.

VI. CONCLUSION

The presented work aims to optimize DEM modeling in terms of modeling accuracy and modeling speed. This is par-

TABLE VI
MODELING SPEED COMPARISON WITH DIFFERENT METHODS AND DIFFERENT TEST SETS

Test sets	T_{with_SVM}	T_{direct}	$\frac{T_{direct} - T_{with_SVM}}{T_{direct}}$ (%)
	(unit: min)	(unit: min)	
5 Positive + 15 Negative	874.3	911.6	4.09
7 Positive + 13 Negative	958.4	1055.2	9.17
10 Positive + 10 Negative	1085.4	1261.2	13.94
15 Positive + 5 Negative	1294.5	1608.1	19.50

ticularly desired when DEM modeling is based on a large number of terrain tiles covering a large area with different landforms. The main contribution of our work is the effective approach for selection of appropriate terrain modeling methods in forming a DEM. This approach achieves a balance between modeling accuracy and modeling speed. In the development, a terrain complexity index is defined to represent a terrain's complexity. Combined with this index and the terrain elevation range, SVMs classify terrain surfaces into either complex or moderate. The classification result recommends a terrain modeling method for a given data set in association with its required modeling accuracy. LOLA data have been used in our experiments to prove that the designed approach is effective. The statistical results have demonstrated that the average accuracy of SVMs for the testing data set is $84.3\% \pm 0.9\%$ for terrain type classification. For various ratios of terrain types in the selected data set, the DEM modeling speed increases up to 19.5% with given DEM accuracy. The experiments focus on two DEM modeling methods, i.e., grid-based and TIN. It is possible to extend the work to more DEM modeling methods aiming for balancing the modeling accuracy and the modeling speed.

REFERENCES

- [1] J. P. Wilson, "Digital terrain modeling," *Geomorphology*, vol. 137, no. 1, pp. 107–121, Jan. 2012.
- [2] S. Li, R. A. MacMillan, D. A. Lobb, B. G. McConkey, A. Moulin, and W. R. Fraser, "LiDAR DEM error analyses and topographic depression identification in a hummocky landscape in the prairie region of Canada," *Geomorphology*, vol. 129, no. 3/4, pp. 263–275, Jun. 2011.
- [3] A. Wehr and U. Lohr, "Airborne laser scanning—An introduction and overview," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, no. 2-3, pp. 68–82, Jul. 1999.
- [4] H. Ebner, W. Reinhardt, and R. Höler, "Generation, management and utilization of high fidelity digital terrain models," *Int. Arch. Photogramm. Remote Sens.*, vol. 27, pp. 556–566, 1988.
- [5] R. Weibel and M. Heller, "Digital terrain modeling," in *Geographic Information Systems: Principles and Applications*, D. J. Maguire, M. F. Goodchild, and D. W. Rhind, Eds. New York: Wiley, 1991, pp. 269–297.
- [6] Q. Zhou, X. Liu, and Y. Sun, "Terrain complexity and uncertainties in grid-based digital terrain analysis," *Int. J. Geogr. Inf. Sci.*, vol. 20, no. 10, pp. 1137–1147, Nov. 2006.
- [7] Y. Chou, P. Liu, and R. J. Dezzani, "Terrain complexity and reduction of topographic data," *J. Geogr. Syst.*, vol. 1, no. 2, pp. 179–198, 1999.
- [8] Q. Zhou and X. Liu, "Error analysis on grid-based slope and aspect algorithms," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 8, pp. 957–962, Aug. 2004.
- [9] C. Chen and T. Yue, "A method of DEM construction and related error analysis," *Comput. Geosci.*, vol. 36, no. 6, pp. 717–725, Jun. 2010.
- [10] Lunar Orbital Data Explorer. [Online]. Available: <http://ode.rsl.wustl.edu/moon/index.aspx>