Number of Neighbors and Cloud Point Density as Factors Controlling the Quality of DEMs Generated from Airborne LiDAR Data Using Three Interpolators

Alaa A. Elghazouly, Fahmy F. Asal, Mohamed I. Doma

Abstract— Accurate and high spatial resolution digital elevation models (DEMs) are in increasing demand for a growing number of mapping and GIS tasks related to applications such as forest management, urban planning, bird population modelling, ice sheet mapping, flood control, road design, etc. Airborne Light Detection and Ranging (LiDAR) has become the preferred technology for digital elevation data acquisition in a wide range of applications. DEMs quality relies on the quality of elevation data and the process of converting discrete elevation points to a continuous surface represented by a DEM through an interpolation operation. LiDAR technology provides high-density and high-accuracy three-dimensional terrain point data acquisition, however, the quality of the DEM generated from LiDAR is affected directly by the interpolation process. In this study, the Ordinary Kriging (OK), the Local Ploynomial (LP) and the Inverse Distance Weighting (IDW) interpolation methods working under ArcGIS 10.1 were used to study the effect of LiDAR data point density and number of neighbors on the quality of the interpolated DEMs. Statistical tests were applied on the generated DEMs from data files of different point densities. The analysis of results showed that using recommended number of neighbors and elevations point density are time saving and cost effective compared with the use of random values.

Index Terms— Airborne LiDAR, Number of Neighbors, Cloud Point Density.

1 Introduction

Terrain mapping was and still one of the most important concerns of surveyors. The terrain was mapped out using a theodolite and a survey pole that gave relative displacement values from a series of known points. With the advent of cameras and photography in 1839 [1], a new science was born: photogrammetry. Early terrain measurements from cameras were taken of mountaintops to gauge their heights, or conversely, from mountaintops to the ground to map out features. By taking two or more photographs, a stereo-model can be generated which allows for the measurement of features or the ground surface in 3-dimensions. With the invention of the airplane, this survey method was naturally extended to the air and modern airborne photogrammetry began. Traditional methods such as field surveying and photogrammetry can yield high-accuracy terrain data, but they are time consuming and labor-intensive. Moreover, in some situations, for example, in forested areas, it is impossible to use these methods for collecting elevation data. After the invention of laser, LiDAR system appeared [2].

LiDAR is probably the most significant technology introduced in mainstream topographic mapping in the last decade. The main advantage of the technique is that it provides a direct method for 3D data collection. Furthermore, it is highly accurate because of the millimeter- and centimeter-level laser ranging accuracy and

---

- Alaa A. Elghazouly, Associate Lecturer, Faculty of Engineering, Menofia University, Egypt, PH+201001721508. E-mail: alaa_elghazouly@sh-eng.menofia.edu.eg.
- Fahmy F. Asal, Associate Prof., Faculty of Engineering, Menofia University, Egypt, PH+201003300435. E-mail: fahmy_asal@hotmail.com.
- Mohamed I. Doma, Associate Prof., Faculty of Engineering, Menofia University, Egypt, PH-+201025181167. E-mail: zeyad1612002@yahoo.com.

precise sensor platform orientation supported by an integrated Position and Orientation System (POS). Unlike the traditional photogrammetric methods, LiDAR directly collects an accurately georeferenced set of dense point clouds, which can be almost directly used in basic applications. However, the full exploitation of LiDAR's potentials and capabilities challenges for new data processing methods that are fundamentally different from the ones used in traditional photogrammetry [3].

Airborne Laser Scanning (ALS) observations can havemuch more dense point spacing than is typically derived from photogrammetry, with current systems abilities exceeding 1 point /meter. Generating accurate Digital Elevation Models (DEMs) from LiDAR data are in increasing demand. As LiDAR system provides high accurate elevation data, the quality of DEMs depends on the interpolation process. Therefore, it is uneconomic to generate low quality DEMs from LiDAR high-density data because of the interpolation process. There are many interpolation methods to interpolate points such as Inverse Distance Weighting (IDW), Spline, Kriging, and Trend. IDW, LP and OK methods will be used in this test. Mathematically, IDW method defined by the following formula (1) [4]:

(1)

Where: is the estimated value of the point elevation at an un-sampled location, are the data points elevations, is the distance between each data point to the point at an un-sampled location, p is the power and n is the number of neighbors.

Local polynomial fitted to a local subset defined by a window. The size of this window needs to be large enough for a reasonable number of data points to be included in the process. One further adjustment is made to this procedure a measure of distance-based weighting is included, so the least squares model is in fact a weighted least squares fit. The weights are computed using a power function of distance as a fraction of the window size. The simplest case is where the moving window is a circle with radius R. If the distance between grid point (x_i,y_i) and a data point (x,y) within the circle is denoted d_i, p is a user definable power then the weight w_i is given by equation (2) and the least squares procedure then involves minimizing the expression given by formula (3)]5[.

(2)

(3)

On the other hand, Kriging takes into account both the distance and the degree of variation between sampling data. Ordinary, Simple, Universal, Probability, Indicator, and Disjunctive Kriging are available in the Arc GIS Geostatistical Analyst. In this research, Ordinary Kriging was used. Ordinary Kriging (OK) focuses on the spatially correlated component and uses the fitted semivariogram, a diagram relating the semivariance to the distance between sample points used in Kriging, directly for interpolation. The estimator of ordinary Kriging is given by equation (4)]5[:

(4)

Where: is the estimate value of elevation at x_0, is the measure value at the xi and is the weight assigned for the residual of $z(x_i)$.

2 Study area and data sources

The studied site is a forestry area sited in Gilmer County (38° 55′ 12″ N, 80° 51′ 0″ W), West Virginia, USA with an area of 1 km by 1 Km. By view the test site through Google Earth (Fig. 1), it was found that the test site contains variety of features including forests of tall trees in addition to residential area of small rural buildings and some roads. The maximum elevation of data is 451.72 m and the minimum elevation is 308.39 m.

Fig. 1. Gilmer County, West Virginia [6].

The test data is a text file containing the last return data of the scanned points using LiDAR system. This data contains the Elevation, North, East and Signature of each scanned point. This file was downloaded from the Gilmer County, West Virginia, USA website: http://www.wvview.org/data/lidar/gilmerlidar.htm. Gilmer County LiDAR data was collected between March 25 and April 7, 2004. The data was acquired by Airborne 1 Corporation, on behalf of the Canaan Valley Institute. Only first and last returns data are available. Fig. 2 shows a plan view of the scanned data points.

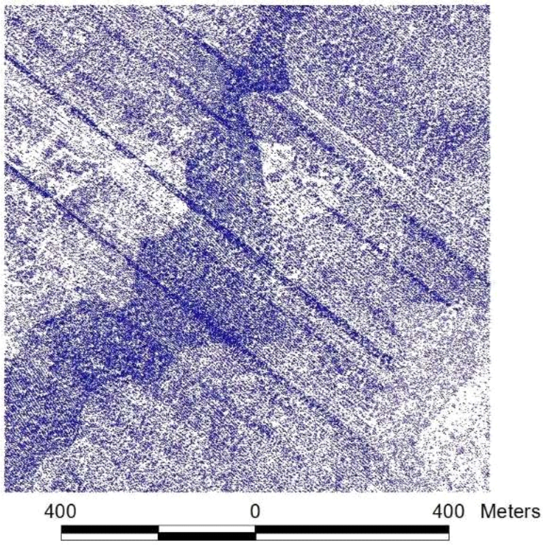

Fig. 2. Layout showing the training dataset.

3 Methodology

In this research, ArcGIS 10.1 with the spatial analysis and the 3D analyst extensions will be used in creating the DEMs from airborne LiDAR data. There are some factors that their values must be entered into the program before interpolation process such as grid cell size, number of neighbors and power. Number of neighbors factor is chosen to study its effect on the DEMs quality. Besides, it is required to determine the number of neighbors that gives the best quality of generated DEMs. LiDAR data point density is also chosen as a factor controlling the quality of DEMs. Not only data point density controls the quality of DEMs but also it controls the cost of LiDAR data. Therefore, it is cost effective to find a relation between the accuracy of interpolated DEMs and data point density.

The sequence of the methodology can be summarized in the following points:

- Converting the data file to text (Tab-delimited) format using Excel program so as to ArcGIS can read it.
- Extracting three files with different cloud point density (80%, 60% and 40% density) from the test file (100% density) using Access program. The test file of last return points has 368052 points and point density 0.368 point/m2 (100% density). The extraction was done by uniformly remove of points after arrange them from smaller to greater according to their North and East values. An example, to obtain 80%-point density two point of each successive ten points were removed. Now there are four test files.
- Dividing the test data into two groups. The first group is the test point dataset, which consists of 3% of number of points of the original files points distributed over the entire test area (Fig. 3) and the second group (Training data set) is the remaining points (Fig. 2).

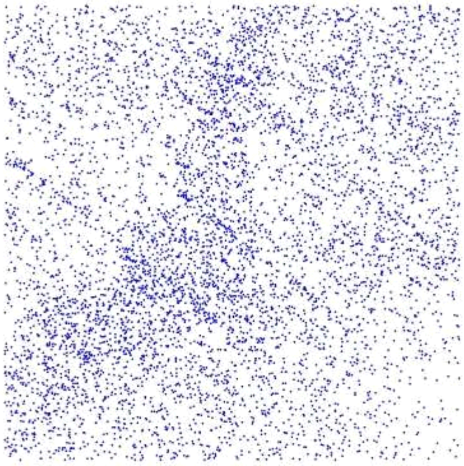

Fig. 3. Test points distributed over the test area.

- For each test file of the four files, Arc GIS 10.1 was used to create 2D views of DEMs using number of neighbors starts from 4 to 32 by step equal 8. Only number of neighbors allowed to change while other factors remain constant (cell size=0.25m and power=2 density=100%). Finally, for each test file there are 4 created DEMs with different number of neighbors.
- Also, different DEMs were created from different point density datasets while the other factors remains constant (cell size=0.25m, power=2 and number of neighbors=12).
- Validation test was used to compare the interpolated elevations of the test checkout points with their real scanned elevations and apply statistical tests to find the standard errors that represent the accuracy of the generated DEMs.

4 Results and analysis

Table 1 shows the summary of the standard errors of DEMs generated from reduced last return LiDAR data using checkpoints and fig. 4 shows the Graphic representation of standard errors results.

From table 1 and fig. 4, the following results have been obtained:

For Ordinary Kriging (OK) interpolation method, the value of standard error increased slightly (1 cm) when LiDAR data density decreased from 100% to 60% and from 60% to 40% density, the standard error increased by 1.5 cm. Local Polynominal (LP) interpolation method standard error also increased by 1.2 cm by decreasing density from 100% to 80%, 1.5 cm from 80% to 60% density and by 2.8 cm from 60% to 40% density. From the Inverse Distance Weighting (IDW) interpolation method, standard error value increased by 1 cm when density decreased from 100% to 80%, 2.9 cm from 80% to 60% and 4.5 cm from 60% to 40%. Fig. 4 shows that OK has the least standard error value and then LP and finally IDW. In addition, it is noticed that IDW is more sensitive to decreasing in cloud point density forwarded by LP and finally OK.

Table 1

Statistical values of validation errors of DEMs generated from different cloud point density using the three interpolation methods

Density	Interpolation method	Statistical Property (m)	
		Range	St. error
100%	OK	2.219	0.146
	IDW	2.591	0.193
	LP	2.96	0.157
80%	OK	2.735	0.152
	IDW	3.238	0.203
	LP	2.769	0.165
60%	OK	2.148	0.162
	IDW	3.039	0.229
	LP	2.53	0.181
40%	OK	2.571	0.177
	IDW	3.508	0.274
	LP	3.061	0.209

Table 2 shows the summary of the standard errors of DEMs generated from 100%-point density dataset last return LiDAR data using different number of neighbors and fig. 5 shows the Graphic representation of standard errors results.

Table 2

Statistical values of validation errors of DEMs generated using different number of neighbors using the three interpolation methods.

Number of Neighbors	Interpolation method	Statistical Property (m)	
		Range	St. error
4	OK	2.554	0.167
	IDW	2.814	0.218
	LP	20.859	0.255
12	OK	2.219	0.146
	IDW	2.591	0.193
	LP	2.965	0.157
20	OK	2.211	0.145
	IDW	2.900	0.190
	LP	3.095	0.163
32	OK	2.180	0.145
	IDW	3.081	0.196
	LP	3.125	0.169

From table 2 and fig. 5, the following results have been obtained:

The standard error of OK method decreased with increasing number of neighbors from 4 to 12 by 2 cm, after that the standard error value remains constant with increasing number of neighbors from 12 to 32 that means increasing number of neighbors more than 12 is time wasting without any accuracy improvement for OK method. For IDW, the accuracy of DEMs improved by increasing number of neighbors from 4 to 20 after that the accuracy decreased. LP method accuracy improved significantly when number of neighbors increased from 4 to 12 by 10 cm, after number of neighbors 12 the standard error increased by increasing the number of neighbors. The following fg.5 can conclude the relation between number of neighbors and the accuracy of different interpolators. OK seems less sensitive for number of neighbors, IDW accuracy improved by increasing number of neighbors till 20 after that the accuracy decreased and finally LP method accuracy is high sensitive of number of neighbors less than 12 after that the accuracy almost remains constant.

5 Conclusion

“Higher is not necessarily better”, that clearly was noticed from the results. Number of neighbors is an effective factor in IDW interpolation method. The interpolated points affected by increasing number of neighbors. That was seen from results of statistical tests of DEMs generated from last return data files. One of the main aims of this study was obtaining the optimum number of neighbors that achieves the higher possible accuracy of interpolated points. Using number of neighbors higher than the optimum number is time wasting and gives a reverse result by reducing the accuracy as seen from results of DEMs generated from last return data file. It is important before starting any scanning project to know the cloud point density to give the required accuracy of finally generated DEMs. The other objective of this study is made a comparison between the DEMs generated using different interpolation methods, OK method has the lowest standard error value (14.6 cm) but it consumes much more time in processing. IDW and LP methods give satisfactory results of standard errors (15 and 19 cm) and less processing time. The effect of processing time will have a big influence when interpolating a large area (not 1 km^2) with high density of LiDAR points, which become more than one point each square meter. It is important before starting any scanning project to know the cloud point density to give the required accuracy of finally generated DEMs. Choosing the best point density depends on the required accuracy and the topography of the area. Statistical tests were applied on the DEMs generated from the last return data (with 122 m difference between maximum and minimum elevation) showed that reducing point density from 0.368 point/m^2 to 0.147 point/m^2 reduced the accuracy of DEMs using OK as an example from 14.6 cm to 17.7 cm. All interpolators affected by decreasing point density but without the same level. OK method seems to be less sensitive for point density as its accuracy reduced by 1 cm for each point density reduction. The accuracy of LP method began to deteriorate after density 40% (0.147 point/m^2) before that density the accuracy reduced equally by 1 cm each point density reduction. IDW method seems to be the most sensitive method to point density that its accuracy deteriorates rapidly after point density 60% (0.22 point/m^2). It can be concluded that, it is not the point density (0.368 point/m^2) that IDW is the best interpolation method in accuracy and time of processing. Increasing LiDAR point density will make local interpolators such as IDW and LP perform better than OK from the aspect of accuracy and processing
time. The elevations of interpolated points have been affected by increasing number of neighbors. That was noticed from results of statistical tests of DEMs generated from last return training data file. The accuracy of OK method is not affected by increasing number of neighbors more than 12, it is only reduced when using number of neighbors 4. For LP, the accuracy reduced when using number of neighbors not equal to 12. IDW accuracy needed number of neighbors more than 12 in order to give the best accuracy. For each interpolation method there is an optimum number of neighbors-depending on topography of land and data point density-must be used to give the highest possible accuracy.

It is important to note here that the process of generating the different DEMs from the training LiDAR data has been undertaken using the default value settings of the different interpolation paratmeters set by the processing softwere (ArcGIS 10.1); this has been for the three tested methods; OK, LP and IDW. In tis case, further investigations through changing the values of the different interpolation parameters of each method reaching to the optimal values of these parameters can give better understanding of the effectiveness of each method in the creation of digital elevation models from LiDAR data. Furthermore, in this case more accurate and fair comparison between the different interpolation methods in the creation of DEMs from LiDAR data can be decided taking into account the time requirments and the cost effectiveness.

6 Referenceso

- Kraus, K. “Photogrammetry Volume 1: Fundamentals and Standard Processes”, English Edition. Köln: Dümmler, 2000.
- Morin K. W., 2002, “Calibration of Airborne Laser Scanners”, PhD thesis, Department of Geomatics Engineering, University of Calgary.
- Shan J. and Toth C.K., 2008, “Topographic Laser Ranging and Scanning Principles and Processing”. ISBN-13: 978-1-4200-5142-1.
- Burrough, P. and McDonnell, R., 1998. “Principles of Geographical Information Systems”, Oxford.
- Eldrandaly, K. A. and Abu-Zaid, M. S., 2011, “Comparison of Six GIS-Based Spatial Interpolation Methods for Estimating Air Temperature in Western Saudi Arabia”, *Journal of Environmental Informatics* 18(1) 38-45.
- Google Earth, 2015.