Computers & Geosciences 64 (2014) 61–71

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Reduction of deformations of the digital terrain model by merging interpolation algorithms

Dariusz Gosciewski *

University of Warmia and Mazury in Olsztyn, Faculty of Geodesy and Land Management, Institute of Geodesy, Heweliusza 12, 10-957 Olsztyn, Poland

A R T I C L E I N F O

Article history:
Received 14 January 2013
Received in revised form
2 December 2013
Accepted 2 December 2013
Available online 17 December 2013

Keywords:
GIS
Digital terrain model
Regular GRID structure
Interpolation algorithms
Large observation sets

A B S T R A C T

Majority of contemporary spatial information systems allow generation and visualization of the digital terrain model based on GRID type of the regular squares network. The values at nodal points for such a GRID are in most cases computed by one of the many available interpolation methods on the basis of dispersed measurement points. The commonly used solutions allow in most cases application of a selected interpolation method in a global way within the entire analyzed area, which does not secure the same accuracy of surface generated. The paper describes generation of a digital terrain model by applying a combination of interpolation methods. The criterion of method choice was dependent on the dispersion of measurement points around the GRID node. This solution allows successive complementing the resultant set of computed set with values determined with a specific error and as a result improves the accuracy of the model generated.

1. Introduction

The development dynamics of contemporary economy causes the demands formulated for spatial information systems to increase continually (Goodchild and Longley, 2005; Harmon and Anderson, 2003; O'Sullivan and Unwin, 2003). Despite continuous development in techniques for gathering, warehousing and analyzing the information originating from various sources there are still limitations in the dynamics of processing it (Arctur and Zeiler, 2004; Egenhofer and Kuhn, 2005; Masser, 2005; Shi et al., 2004). Securing continuous monitoring of dynamic physical phenomena also requires a maximally automated and optimized process for processing the information obtained. Processing large volumes of data in real time is becoming one of the particularly important problems (Pigozzi, 2004; Shary et al., 2002; Wechsler, 2003). Currently, integrated measurement systems such a aerial scanning, laser measurement stations or multibeam echo sounders allow obtaining a very large volume of measurement data (megadata) over a relatively short time (Aruga et al., 2005; Axelsson, 2000; Wack and Wimmer, 2002). Such information, however, is in most cases unsuitable for direct use in spatial analyses mainly as a consequence of its volume, random location and density as well as non-continuous character of measurements sets (Gosciewski, 2013; Heuvelink et al., 2006; Worboys and Duckham, 2004; Zhou and Liu, 2004). One of the methods applied to organize information describing the surface (of the terrain, the sea bottom or artificial objects with a continuous surface such as wall surfaces of buildings, dams or mine workings) is to present it in the form of a regular GRID-type node structure (Gosciewski, 2013a; Li et al., 2005; Raaflaub and Collins, 2006). This allows a significant reduction in the volume of data, limits redundancy and allows segregation of sets, which is particularly useful in case of spatial analyses made over time. Implementation of these tasks requires application of interpolation algorithms allowing fast processing of large volumes of information (Carlisle, 2002; Maune, 2001; Heo, 2003; Zhou and Liu, 2004a). As a consequence it is important to analyze algorithms of that type for accuracy of computations generated by them considering at the same time the efficiency of the processing process.

Many different interpolation algorithms can be used to compute the value at the node (Carlisle, 2002; Lepere and Trystram, 2002; Raaflaub and Collins, 2006; Schabenberger and Gotway, 2005; Zhou and Liu, 2004a). They are characterized by different degrees of complexity and different operational efficiencies. When processing large data volumes (LiDAR), a particularly important factor of an algorithm's operation is its computational speed. Classical interpolation methods used widely in commercial applications allow the values at all GRID nodes in a specific area to be determined globally by one selected method.

Because there is unlimited number of morphological surface combinations, it is not possible to point out an existing interpolation algorithm which would unquestionably generate the most accurate model. Moreover, each of the algorithms has a number of processing parameters. The analysis of all parameters on all

* Tel.: +48 89 5233582.
E-mail address: chillis@uwm.edu.pl

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.12.005

surface combinations is impossible, but some features of the algorithms (e.g. their complexity interrelated with computational speed) can be used for combining them into a sequence and raise processing efficiency. The use of one interpolation method for the entire area in the case of large number of measurement points is not always justified in terms of accuracy and economic efficiency. Although the application of simple, fast algorithms obtains less accurate interpolation results, the selection of precise (and thus more complex) algorithms is not very efficient for a large volume of data. It is therefore worth testing a solution which combines several algorithms operating in the same area.

Several variants of systems of several different algorithms were studied to analyze the problem. In most cases, the combination of algorithms proved more economical and gave more accurate results than the application of one algorithm for the whole area. A system of three different algorithms: from the simplest (fastest) to one with medium complexity to the most complex (slowest) proved most accurate and most economically advantageous (a relatively short processing time).

The presented example shows the general methodology of the procedure during the use of a combination of several algorithms. It consists of gradual computation of values at selected nodes at successive stages of processing by individual algorithms (from the fastest to the slowest). At each stage, only the nodes which fulfill the assumed accuracy condition remain in the set of nodes determined by the given algorithm and the other nodes are eliminated from the set. In successive stages, increasingly accurate algorithms perform computations on diminishing sets because interpolation no longer includes the nodes determined by the previous methods. The gradual decrease in the number of nodes for interpolation at individual stages leads to the situation where increasingly complex algorithms have less and less data to process. Due to this, the whole computational process shortens while maintaining the required accuracy.

2. Test model development

Meaningful analysis and comparison of accuracy of various interpolation algorithms should be assumed using the same base of measurement points. For that purpose the standard base of points situated on a theoretical surface model was established. Using the function of two variables (1) the mathematical surface presented in Fig. 1 was generated.

$$f(x,y) = \sin(xy + e^y) + \cos(xy + \sqrt{x}) - x - y \quad x \in <2;5>, \quad y \in <-1;2> \tag{1}$$

Any number of points, which will be equivalents of the measurement points, can be generated on the basis of their function. Their density can correspond to the characteristics of modern LiDAR systems, where the number of measurement points per unit of area reaches 20 pts./m^2. The theoretical model allows to set any measurement point density proportion and GRID resolution. GRID resolution can be higher or lower than the survey resolution. A high number of measurement points per one grid node causes interpolation algorithms to reach similar accuracies. However, such a large number of points is not always available. In the case of filtering the measurement set and discarding the points not lying on the surveyed terrain (reflected from trees, poles, building roofs, etc.), the data volume decreases considerably. It is therefore worth testing the capabilities of specific interpolation algorithms in the situation of a limited number of measurement points. This allows the shortcomings of individual algorithms to be detected and select their sequence, enabling improvement in the accuracy of the generated surface model.

Within the given range of independent variables x and y ,10,000 points being equivalent to measurement points were generated in a random process (Fig. 2a). Next, the entire standard area was rescaled and moved in such a way that it formed a square with side length of ca. 450 m, where: x min = 10,300.01 m, x max = 10,749.97 m, y min = 9850.02 m, y max = 10,299.93 m, z min = 116.22 m, and z max = 161.15 m. Using the same function (1) the theoretical GRID of nodes forming the GRID structure consisting of 2116 points distributed in corners of base squares 10 m × 10 m was also generated (Fig. 2c). For the purpose of securing a sufficient number of measurement points around each node, 180 edge nodes were excluded from further analysis. The final theoretical GRID of squares contained 1936 points (44 × 44), with the coordinates of the starting point: x = 10,310.00, and y = 9860.00. Next, applying selected interpolation algorithms, on the basis of measurement points generated earlier, the values at nodal points were computed. This allowed generating interpolated surfaces consisting of the practical GRID of nodes with the same x and y coordinates as the theoretical GRID. The RMS coefficient, the value of which was computed for each interpolated surface on the basis of formula (2) was used for comparison of match between the interpolated surface and the standard surface (Cressie and Pavlicova, 2002; Maune, 2001; Schabenberger and Gotway, 2005).

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(x_i, y_i) - h_i)^2} \tag{2}$$

where:

- $f(x, y)$ – value of function (1) in the theoretical nodal point with coordinates x y,

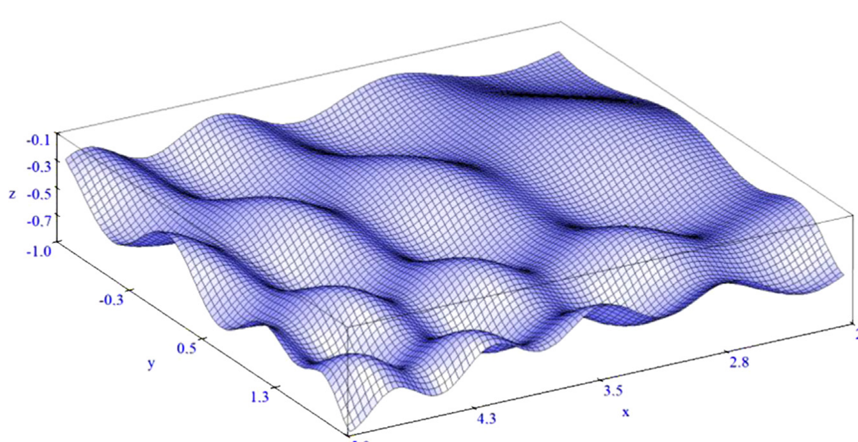

Fig. 1. Graph of 3D function.