

LIDAR DATA POINTS FILTERING USING ARCGIS' 3D AND SPATIAL ANALYST.

Juliano Kersting
Ana Paula B. Kersting

ABSTRACT

One of the main advantages of the airborne laser scanner systems is the high degree of detail that a portion of the land can be mapped. This overdetailed description is caused by the high number of acquired points, which makes it easier to identify objects and modeling the topography. However, the large amount of collected points becomes redundant in plain regions, where fewer points are needed to describe the surface. An algorithm aimed primarily at the reduction of the number of points within a TIN model produced using LIDAR data was implemented in C# language using ArcObjects and both 3D and Spatial Analyst ArcGIS extensions. The method is based on the faces of the triangulation, where the redundant points are eliminated by a neighborhood vertex importance analysis. The obtained results with different thresholds are presented and map algebra calculations on raster created with two generalized subsets are used for evaluation.

Keywords: LIDAR, Triangulated Irregular Network, TIN, Digital Surface Model, Filter.

1. INTRODUCTION

The representation of a surface can be generally seen as a problem of modeling in $2^{1/2}$ dimensions, where a function of two variables $z = f(x, y)$ expresses the surface elevation at the point (x, y) in the Euclidean plan (Burrough 1986). Then, any line, parallel to z axis, penetrates in the surface only once. This representation is appropriate for the most of land types, although some features such as caves and faults are excluded. The model in question can be associated to the points located in the terrain surface, when this is called Digital Terrain Model – DTM, as well as to the surface of all objects in the region, including the terrain and other objects as vegetation and buildings. In this in case, the model is called Digital Surface Model - DSM. The choice of the model used depends on some factors such as the nature of the input data, the available domain of the application and computational resources. The most common is the one that uses a regular mesh to model the region. This representation also is known as Raster. A great disadvantage of these models is its space invariability, since the structure does not adjust to the irregularities of the surface. This can result a great redundancy in the data, especially in plain areas, where minimum topographical information is needed.

As an alternative, there are the Irregular Triangulated Networks - TIN, that can be considered as approaches of topographical surfaces through a set of contiguous not overlapped triangular faces, generated from a finite set of points (Chen & Guevara 1987). There are many advantages associated to the use of the TIN model. The surface data can be irregularly distributed in the space. Moreover, features can be incorporated the model. For example, vertexes in a TIN can describe nodes such as depressions, peaks or ways, while edges can represent linear features such as breaking lines, ridges or canals. They are difficult to acquire, due mainly to the massive data and some problems concerning land analysis and also the lack of available computational resources, although TIN models have been widely used for its efficiency in the storage of irregularly distributed elevation points.

The airborne laser mapping, called LIDAR (**L**ight **D**etection **A**nd **R**anging), provides fast acquisition of a great volume of information on the variation of the surface region through a high density of three-dimensional points, allowing a great quality representation of the land surface. However, the large amount of measurements generates redundancy in plain regions, where a small density of points is needed in order to describe the surface. By providing a high degree of detail to a TIN representation, this technology demands a great storage capacity and computational resources. Then, to be effective, a model must be well weighted regarding the requirements of resolution and storage space.

To overcome computational restriction, a method of generalization of the digital surface model produced by the laser scanning technology, based on the reduction of redundant points using irregular nets of triangulation is described in this paper. The basic principles of LIDAR technology are presented and an analysis of its data accomplished in order to show the redundant points in a mapped area. Related work for the solution of this kind of problem is presented and the results of the proposed method discussed.

2. THE LIDAR SYSTEM

Functionally, LIDAR is the result of the integration of three technologies in a system capable of data acquisition for the production of Digital Surface Models (DSM). These technologies are: a Laser Scanning and Ranging System, a Global Positioning System (GPS) and an Inertial Navigation System – (INS) also known as Inertial Measurement Unit (IMU). Combined, they provide the positions where the laser beams touch the surface with high precision. The Figure 1 illustrates the system components. Laser pulses are highly accurate in their ranging capabilities, providing distances with accuracy to within a couple of centimeters. The accuracy limitations of the LIDAR technology are due primarily to the GPS and IMU systems (Wehr & Lohr 1999).

A LIDAR system combines a single narrow-beam laser with a receiver system. The system produces optic pulses in direction to the surface, where they are reflected off and returned to the receiver. The receiver accurately measures the travel time of the pulse from its start to its return. As the pulse travels at the speed of the light, the receiver senses the return pulse before the next pulse is sent out. Since the speed of the light is known, the travel time of the pulse can be converted into distance. Combining this distance with the angle of emission of the laser pulse with the GPS position of the laser scanner and the orientation of the system provided by the IMU, the value of x, y, and z positions can be calculated for each emitted laser pulse.

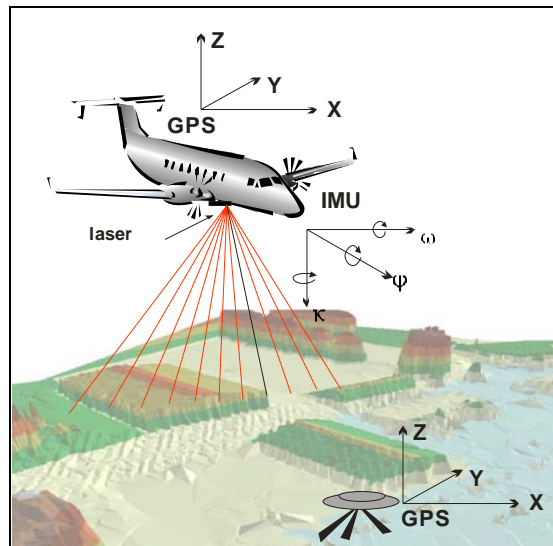


Figure 1- LIDAR system components.

The scanning of a surface is transversal to the direction of flight. The width of the strip or "swath" covered by the ranges depends on the scan angle of the laser ranging system and the airplane height (Figure 2).

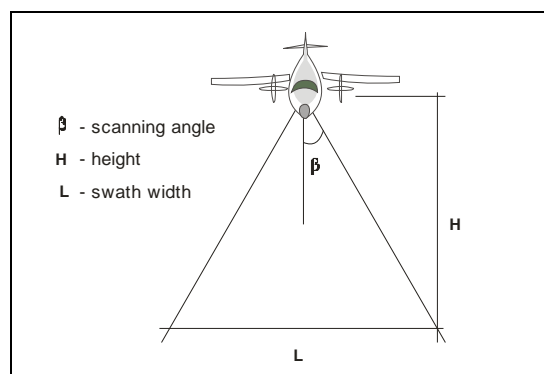


Figure 2 – Swath width.

The point density acquired depends on the system pulse repetition rate, the flying height, the scanning angle, the flying speed and the scanning rate. Thus a higher density of points is achieved when flying lower, with a small scan angle and at a lower speed for a given pulse repetition rate.

The resulting product of the laser scanning is an ASCII text file containing three-dimensional points describing the land surface and objects for use in DSM generation. The data are distributed in an irregular way and due to its high density, may become redundant, especially in plain regions. The representation structure that better preserves the original information of the survey is the TIN model, but in order to reduce the data volume, regular grid structure is frequently employed. In both cases, the generation of the model requires the analysis of the whole data set, becoming difficult and time consuming, considering the great volume of data. For this reason algorithms aiming the reduction of the data set, preserving surface information are desired.

3. PREVIOUS WORK

The main idea concerning generalized terrain models (Chen & Tobler 1986, Of Berg & Dobrindt 1998, Floriani et al. 1996, Voigtmann et al. 1987) is that the representation of a surface in an arbitrary level of detail can be carried through by the insertion of significant points in a rough model or by the removal of less significant points of a detailed model. Most of surface simplification methods found in literature can be classified as methods of refinement and decimation (Pedrini - 2000). The refinement method starts with a coarse approach of the surface and repeatedly adding points to the triangulation until the model satisfies a specified approaching criterion. The decimation method initiates with the triangulation model containing the entire set of points, simplifying it iteratively until the approaching criterion is satisfied. The main objective of a decimation algorithm is to reduce the number of triangles in a mesh, preserving the main features at the best definition. The algorithm used in this work fits in this last group.

Most of decimation algorithms can be classified in three categories, according to the geometric entities used for the removal. The main groups are:

- A. Nodes decimation (Fig. 3.A). That was adopted in this work and some examples found in literature are presented to follow.
- B. Edges decimation (Fig. 3.B). In this group are the methods argued at Hoppe et al. (1993), also the one at Hoppe (1996), and the method of Guézic & Hummel (1995).
- C. Decimation of Triangles (Fig 3.C), considered by Scarlatos & Pavlidis (1990);

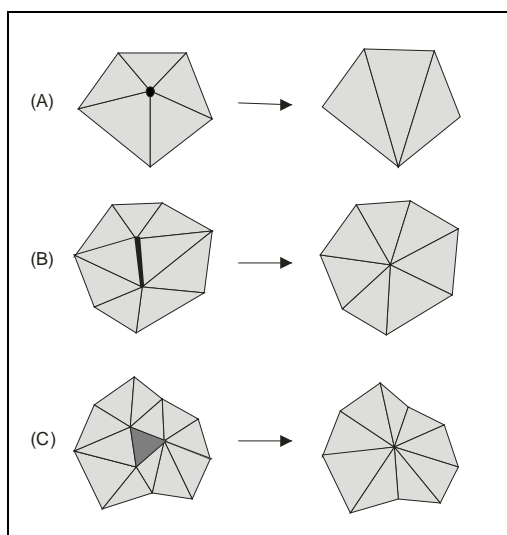


Figure 3 – Decimation Categories: (A) by Vertex; (B) by Nodes; (C) by Triangles

Lee (1989) considered a method of terrain simplification called drop heuristic. The algorithm uses an iterative vertex decimation approach, removing a point on each step. An initial triangulation is created adding a line that connects diagonal point to each 2x2 neighborhood. The error, measured as the

differences in the elevation between the two surfaces is calculated for each remaining point in the triangulation, and the point with the inferior error is excluded.

A general decimation algorithm is presented by Schroeder et al. (1992). The algorithm executes multiple times on an existing triangulation, removing vertex until a specified value of error is reached. The error on the vertex is calculated based on the value of the distance from the point to the average plan of the neighbor vertex.

The simplification can also be done through the triangulation of a new set of vertex in replacement for the original triangulation, as considered by Turk (1992). In this approach, an iterative procedure of point repulsion is used to distribute the new set of points over the surface, concentrating more points in the region with higher curvature. Then the original points are removed one by one, resulting in a triangulation that preserves the topology of the original surface.

Schroder F. & Robbath (1994) described a decimation algorithm where the importance of a vertex is evaluated in accordance with the measure of roughness of the surface in that vertex. A vertex is removed just in case it does not have significant change in the general representation and the area surrounding the removed point is re-triangulated.

A similar approach was adopted in this work where the criterion for the vertex selection is crucial during the generalization process since this can establish the quality and precision of the resultant mesh. The technique used in this work, the most used for precision evaluation, is known as L8 rule that is based on the maximum difference between the original elevation data and the generalized surface. The difference corresponds to a measure of local error. Another common approach is the L2 rule, that allows a measure of average shunting between the original and the approached model.

4. METHODOLOGY

The main idea of the considered method is to produce a representation of low resolution that enables the extraction of an approach model for a given tolerance, meeting the condition, through a compact and fast process. The structure of adopted data in this in case is TIN structure, for considering that the most information of the original data can be preserved, especially in the edges of objects. The first stage of method is the generation of an initial mesh using Delaunay triangulation to ensure that only triangles with consistent geometry are created from the three-dimensional points. This initial triangulation is generalized iteratively through the removal of low importance vertex, ever preserving the topological characteristics of the considered model. The decimation algorithm is iterative and generates a simplified grid derived from the removal of the redundant points. The criterion for removing a vertex v is calculated as the average of the normal n_i of the triangles surfaces in the surroundings of v weighed by the values of the areas A .

$$\alpha_{\max} = \max \left(\arccos \frac{\bar{n}_{av} \cdot \bar{n}_i}{|\bar{n}_{av}| \cdot |\bar{n}_i|} \right) \quad (1)$$

where

$$\bar{n}_{av} = \frac{\sum \bar{n}_i \cdot A_i}{\sum A_i} \quad \text{and} \quad 0 \leq \bar{n}_i \leq \pi \quad (2)$$

The relevance of the vertex is weighed in function of the local deviation of the surface, described by the slope of neighbor triangles where the point is a shared node (Figure 4).

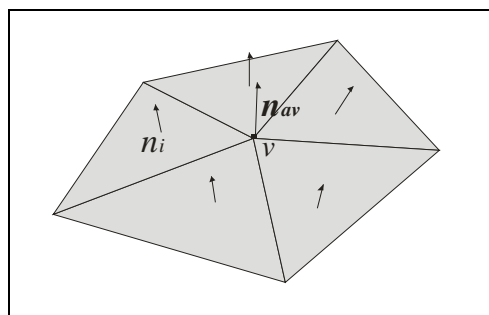


Figure 4 –Vertex Removing Criteria

For each vertex, the gradient of the adjacent triangles is calculated and the average gradient of the region it is estimated. The gradient variation of the region forming by the plans is analyzed by comparing each value to the average value. If discrepant to the average values are verified, the region is considered heterogeneous and the relevant vertex is preserved. Otherwise, the region is considered uniform and the point is redundant and is removed from the vertex collection. The vertex v will be removed if the value of the highest angle between the average normal n_{av} and the normal of the remaining triangles is smaller than a pre-defined threshold. Considering that the process generates new triangles, the relevance of each vertex must be reevaluated until all non important vertexes are eliminated.

5. RESULTS

The considered method was applied in different data sets enabling the evaluation of the accomplished precision and performance in the proposed simplification process. The data derived from digital laser scanning, were generously supplied by the LACTEC - Institute of Technology for the Development. Figure 5 illustrates one of the original TIN used, named *Hospital*, a small area of 13.640 m² enclosing a hospital called Erasto Gaertner (southwest side of the university campus, in Curitiba, Brazil). The model is composed for 17506 points, with average resolution of 1.3 point/m², and elevations between 901.56 m and 925.27 m in a mesh composed by 34986 triangles.

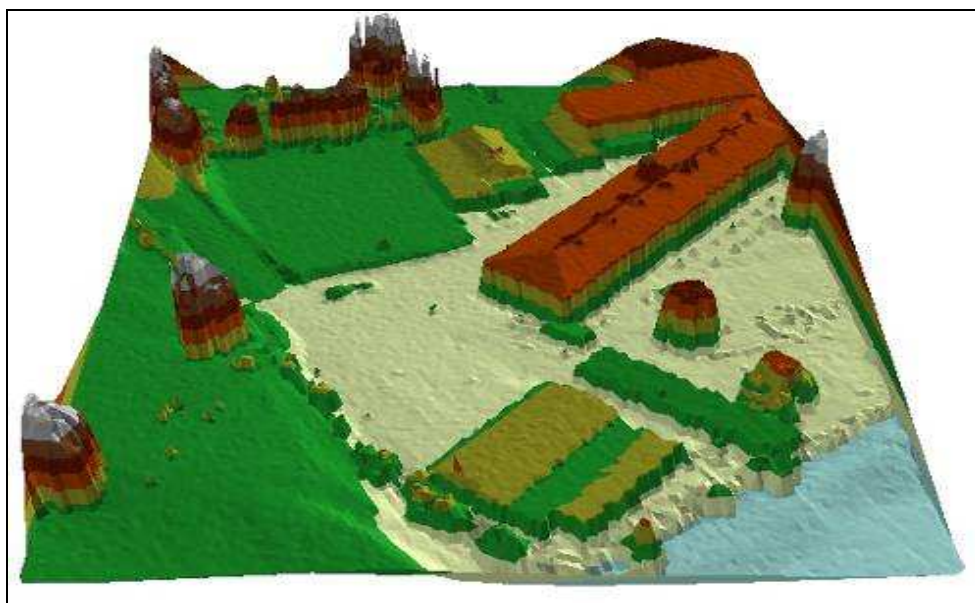


Figure 5 –TIN model of the Hospital (original)

Figure 6, shows the triangulation model named of *The Squares* which encloses a portion of athletics field in an area of 33.850 m² inside the campus, composed of 93904 points, an average resolution of 2.8 points/m², elevation between 900.18 m and 936.04 m, forming a model composed of 187765 triangles. In this area there is an overlapping region formed by two laser covering flight lines, considerably increasing the density of points. This fact becomes more evident in figure 8 and will be later explained in details. Only for visualization purposes the TIN models were symbolized with colors in based on heights, following a color ramp from blue (lower) to white (higher).

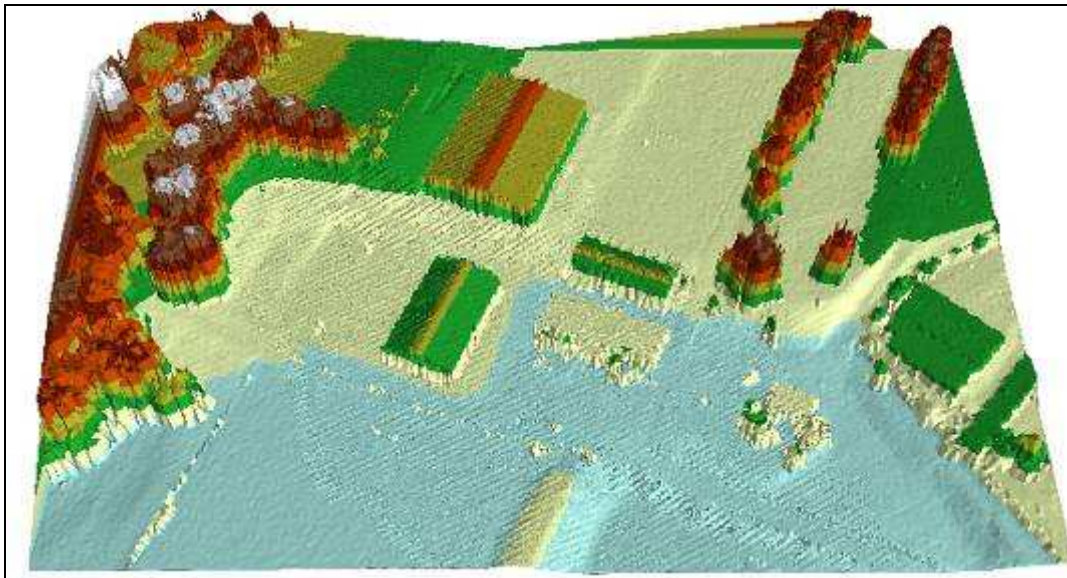


Figure 6 – TIN Model of the Squares (original)

These two models were generalized using threshold values ranging from 5°, 10° and 15° for the points removal criterion, where different results were taken in each of the iterations. Tests were carried out through using a maximum of 20 iterations for each specified criterion range, or until a total of 100% exploitation reached, meaning no more points to be removed. The results are shown in Tables 1 and 2, where the number of points, triangles, points to exclude and the remaining rate of the points are shown.

Table 1 – Obtained results for the Hospital model

ITERATIONS	1	2	3	4	5	10	11	15	16	17	18	19	20
HOSPITAL 5°													
Points	17506	11515	9827	8962	8540	7629	7579	7467	7460	7455	7453	7452	
Triangles	34196	22465	19171	17471	16657	14867	14737	14527	14505	14499	14495	14493	
Points to remove	5991	1688	865	422	335	50	26	7	5	2	1	0	
Remaining rate	65,78%	85,34%	91,20%	95,29%	96,08%	99,34%	99,66%	99,91%	99,93%	99,97%	99,99%	100,00%	
HOSPITAL 10°													
Points	17506	7187	5714	5318	5121	4898	4893	4868	4867				
Triangles	34196	13908	11059	10273	9897	9483	9451	9403	9401				
Points to remove	10319	1473	396	197	110	5	10	1	0				
Remaining rate	41,05%	79,50%	93,07%	96,30%	97,85%	99,90%	99,80%	99,98%	100,00%				
HOSPITAL 15°													
Points	17506	6226	4819	4497	4344	4182	4181						
Triangles	34196	12063	9289	8643	8348	8074	8072						
Points to remove	11280	1407	322	153	76	1	0						
Remaining rate	35,56%	77,40%	93,32%	96,60%	98,25%	99,98%	100,00%						

Table 1 shows that a small increment in point acceptance threshold values drastically reduced the number of remaining points and lowered the redundancy and excessive precision. Also was noticed that in any case the total amount of specified iterations was attained, meaning small processing time and also maximum removal. Table 2 shows the results using the same angles for the second sample area. In this in case the process became slower due to complexity of the mesh formed by two overlapping bands, where it was impossible to reach the complete removal even at 20 iterations.

Table 2 - Results obtained for the Squares model

ITERATIONS	1	2	3	4	5	10	11	15	16	17	18	19	20
SQUARES 5°													
Points	93904	83526	79008	76033	73723	66544	65604	62559	62046	61499	60915	60360	59895
Triangles	88122	71982	66302	63310	60940	55481	54897	52917	52621	52429	51903	51583	51315

Points to remove	10378	4518	2975	2310	1830	940	937	513	547	584	555	465	377
Remaining rate	88,95%	94,59%	96,23%	96,96%	97,52%	98,59%	98,57%	99,18%	99,12%	99,05%	99,09%	99,23%	99,37%
SQUARES 10°													
Points	93904	71176	62267	56295	52014	42372	41320	38032	37381	36756	36190	35647	35115
Triangles	88122	56750	49109	43747	39660	31687	31021	29382	29183	29008	28722	28594	28492
Points to remove	22728	8909	5972	4281	3097	1052	934	651	625	566	543	532	492
Remaining rate	75,80%	87,48%	90,41%	92,40%	94,05%	97,52%	97,74%	98,29%	98,33%	98,46%	98,50%	98,51%	98,60%
SQUARES 15°													
Points	93904	66174	54723	47108	42851	34767	33815	30650	29938	29242	28631	28014	27428
Triangles	88122	52440	42837	34417	29906	24244	23884	22939	22667	22551	22425	22227	22131
Points to remove	27730	11451	7615	4257	2531	952	850	712	696	611	617	586	555
Remaining rate	70,47%	82,70%	86,08%	90,96%	94,09%	97,26%	97,49%	97,68%	97,68%	97,91%	97,84%	97,91%	97,98%

Figure 7 and 9 show successive results from the original Hospital and the Square models using all points using threshold values of 5° , 10° and 15° . Despite the resultant models be composed by only a small percentage of points in relation to the number of points in the original data, the main features in each model had been kept. Was also noticed that the terrain surface suffered high degradation while the boundary of objects were preserved exactly even for the higher angle. This can be considered a loss in certain cases but for most purposes is a great advantage. For example when the goal is to identify the boundary of buildings, the higher angular coefficient is the one that eliminates lesser significant edges. Comparing the results using 5° , 10° and 15° , the ridges of the roofs were removed at higher angles, what makes sense when considering the removal criterion. However for the roof identification and its components, an intermediate angle is more appropriate.

The great advantage shown is the reduction of the data volume. For 5° , the amount of points after the processing is 43% (Hospital) and 63% (Squares) of the original amount. The overhead in the Squares model is associated with the occurrence of the overlapping area that generates a great number of small triangles in all directions and a high discrepancy from the average surface due mainly to the error associated with inertia on mirror motion mainly in higher angles. Also the density of points causes a high frequency of divergent angles in adjacent triangles and reduction of its areas.

To evaluate the difference between the filtered data sets and the original, regular grids were generated by interpolation using both the original point data model and the generalized ones. These new models were compared together and the difference between the original model and the filtered was calculated. The spatial distribution of the differences found in each of the approaches, and a summary statistics for the Hospital area are shown in figure 8 and 10. A similar analysis was lead for another area obtaining similar results.

The difference between the two original grids and the Hospital, with 5° , is null in practical terms. Most of the points have differences smaller then 0.5 meters. As the angle increases, the removal process becomes coarser and the resultant grid diverges more from the original grid. In the extreme case considering 15° , the difference between the grids is significant, since relatively great areas appear having vertical difference values superior to one meter. This fact also can be reinforced by the increasing standard deviation value as the angular threshold increases.

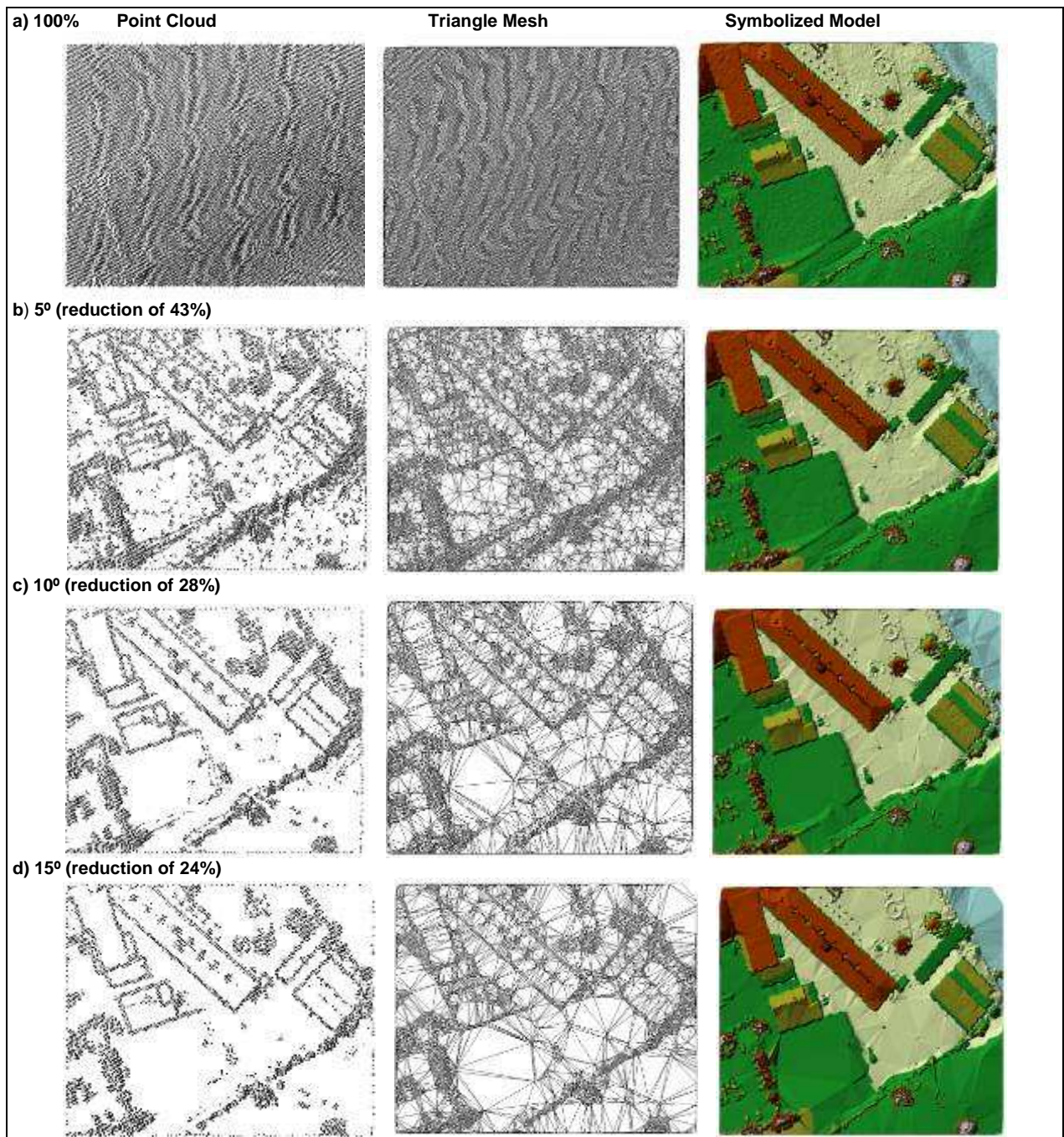


Figure 7 – Simplifications of the Hospital model– a) Original, b) 5°, c)10° e d)15°

Hospital 05°		Hospital 10°		Hospital 15°	
Standard Dev: 0,2017 Mean: -0,0032 Maximum: 5,2058 Minimum: -10,7488	Erro (m) 	Standard Dev: 0,2770 Mean: -0,0137 Maximum: 7,1970 Minimum: -12,4459	Erro (m) 	Standard Dev: 0,3432 Mean: -0,0284 Maximum: 7,1970 Minimum: -12,4459	Erro (m)

Figure 8 – Spatial distribution of errors for the generalized Hospital models.

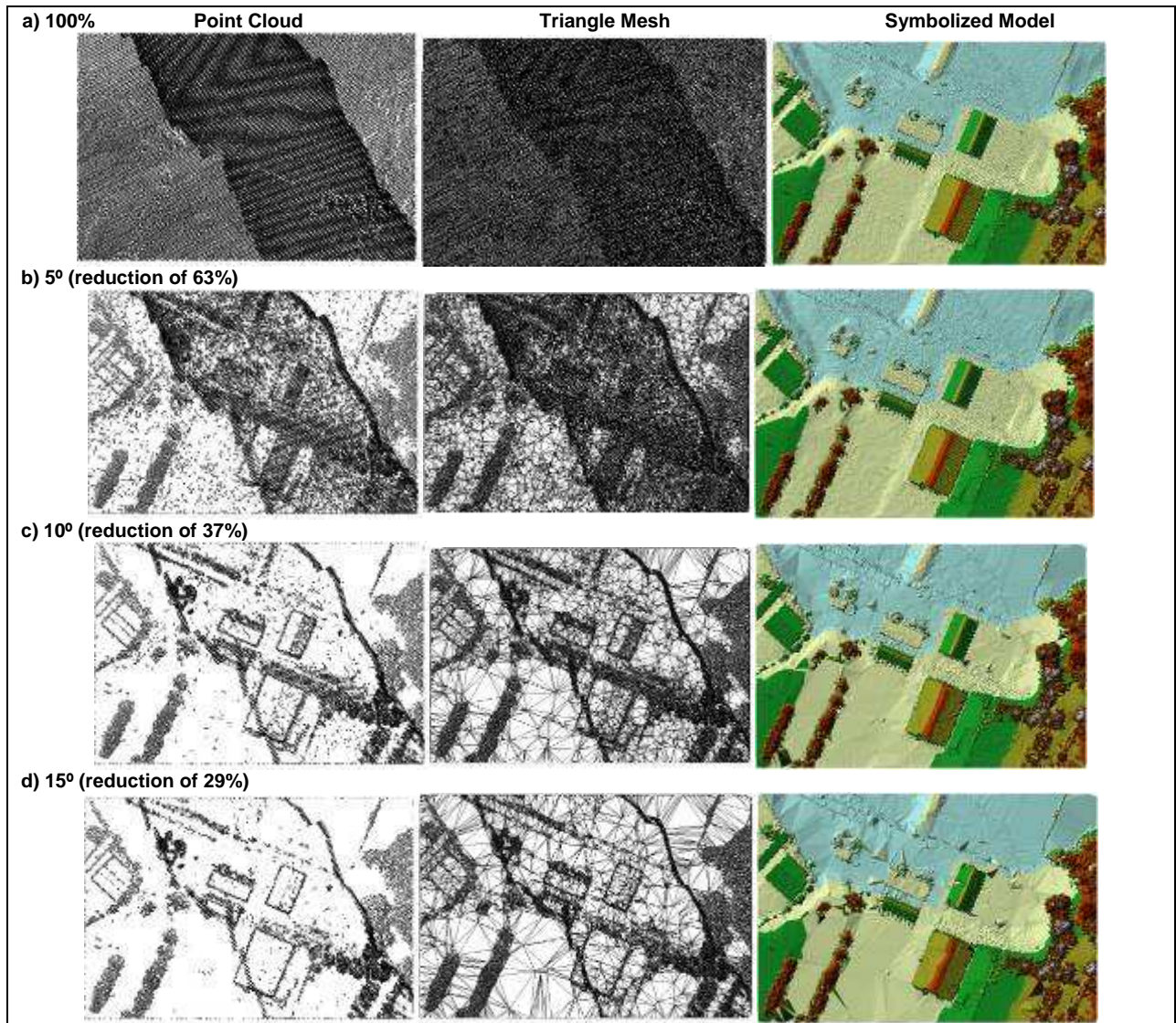


Figure 9 – Simplifications for the Squares model – a) Original, b) 5°, c)10° e d)15°

Squares 05°	Squares 10°	Squares 15°
Standard Dev: 0,2989 Mean: -0,0071 Maximum: 10,7041 Minimum: -12,2661	Standard Dev: 0,4894 Mean: -0,0305 Maximum: 10,7064 Minimum: -12,3545	Standard Dev: 0,6658 Mean: -0,0687 Maximum: 10,6906 Minimum: -13,4128

Figure 10 - Spatial distribution of errors for the generalized Square models.

6. CONCLUSION

The described method in this work allowed the extraction of triangulation models in different levels of detail from a series of generalization operations. This computational process supplied compact models, very smaller in size than the originals and little loss of information, ensuring a good cost benefit relation in terms of both computational efforts and necessary storage space for the resultant models. The tests proved that the elimination of points in plain areas can significantly contribute to the reduction of the data volume without losing information of edges, which are relevant in the surface representation. The presented method depends on the size of the triangles which conditioning the gradient of the surfaces. From an operational point of view, the application using those techniques enables many restricted applications, since the equipment processing capacity was optimized even allowing the extension of the covered area. The reduction can also be applied to generate models with different degrees of generalization that can be used for visualization in some levels. In an initial level, a less detailed model is enough and more economic. That is truth especially for the visualization of three-dimensional models over the Internet, a way that is in constant expansion and whose efficiency of use depends on the volume of data to be transmitted.

ACKNOWLEDGEMENTS

This work was developed with the support of the Federal University of Paraná, Brazil - UFPR. We are also thankful the LACTEC - Institute of Technology for the Development - for the supply of the necessary data and the COPEL - Energy Company of Paraná State, Brazil - for fomenting the research through the incentive to the continuous development, especially in all levels of studies.

REFERENCES

- BURROUGH, P. A. 1986. **Digital Elevation Models: Principles of Geographical Information System for Land Resources Assessment**. *Monographs on Soil and Resources Survey*. Oxford, 1986.
- CHEN Z. & GUEVARA J.A. 1987. **Systematic selection of very important points (VIP) from digital terrain model for constructing triangular irregular networks**. In: Eighth International Symposium on Computer-Assisted Cartography - Auto-Carto 8, Baltimore, Maryland, USA, *Proceedings*, p. 50-56.
- CHEN Z.T. & TOBLER W.R. 1986. **Quadtree representation of digital terrain**, In: Auto-Carto, London, England, *Proceedings*, p. 475-484.
- DE FLORIANI L., MARZANO P., PUPPO E. 1996. **Multiresolution models for topographic surface description**, *The Visual Computer*, 12:317-345.
- GUÉZIEC A. & HUMMEL R. 1995. **Exploiting triangulated surface extraction using tetrahedral decomposition**. *IEEE Transactions on Visualization and Computer Graphics*, 1:328-342.
- HOPPE H., DEROSE T., DUCHAMP T, MCDONALD J., STUETZLE W. 1993. **Mesh optimization**. In: SIGGRAPH'93 Conference, Anaheim, California, USA, *Proceedings*, p. 19-26.
- HOPPE H. 1996. **Progressive meshes**. In: *SIGGRAPH'96 Conference*, New Orleans, Louisiana, USA, *Proceedings*, p. 99-108.
- PEDRINI H. 2000. **An Adaptive Method for Terrain Surface Approximation based on Triangular Meshes**. Rensselaer Polytechnic Institute, Troy, NY, USA, PhD Thesis, 46 p.
- SCARLATOS L. L. & PAVLIDIS T. 1990. **Hierarchical triangulation using terrain features**. *IEEE Visualization'90*, San Francisco, California, USA, *Proceedings*, p. 168-175.
- SCHRODER F. & ROBBACH P. 1994. **Managing the complexity of digital terrain models**. *Computer & Graphics*, 18:775-783.

SCHROEDER W. J., ZARGE J. A., LORENSEN W. E. 1992. **Decimation of triangle meshes.** *Computer Graphics*, 26:65-70.

TURK G. 1992. **Re-tiling polygonal surfaces.** *Computer Graphics*, 26:55-64.

VOIGTMANN A., BECKER L., HINRICHS K. 1997. **A hierarchical model for multiresolution surface reconstruction.** *Graphical Models and Image Processing*, 59:333-348.

WEHR A. & LOHR U. 1999. **Airborne Laser Scanning – An Introduction and Overview.** ISPRS Journal of Photogrammetry & Remote Sensing, pp. 68-82.

AUTHOR INFORMATION

Name: Juliano Kersting

Title: GIS Analyst

Organization: COPEL – Companhia Paranaense de Energia

Address: Rua José Izidoro Biazetto, 158, bloco B, Mossunguê, Curitiba, Paraná, Brasil.

Phone: 55 41 3331 4915

Mobile: 55 41 9996 2600

E-mail: juliano.kersting@copel.com