

Global Surface Model using Space-shuttle Radar Topographic Mission Dataset and Global Indexing Grid

Jan KOLÁŘ, Erik KJEMS, Lars BODUM, Esben Munk SØRENSEN, Denmark

Key words: SRTM, Spatial database, Topographic surface, Global grid, 3-D.

SUMMARY

This article provides methods and results in representing a topographic surface, which were achieved during a research and development project regarding a platform for three-dimensional geographic applications called GRIFINOR. The use of traditional digital elevation/terrain model (DEM) data structures is linked with an original approach devised in the project. The article describes the processing of a dataset from the Space-shuttle Radar Topographic Mission (SRTM) using an original method based on a Global Indexing Grid (GIG). The principle of the GIG method is explained briefly. A freely available global coverage data product of the SRTM has been used for this purpose. This dataset represents an output from contemporary scanning and ranging technologies, which have many properties in common with the state-of-the-art instruments available today. These instruments produce huge amounts of relatively accurate ranging measurements at extremely high rates. This fact however puts the measurement technology ahead of the processing methods, and data management solutions seem to be unable to fully exploit the potential of the measurement technology. This article identifies some sources of this problem and proposes a solution. This article also deals with the flaws of the scanning method used during the space-shuttle mission. A great number of void data were recorded and should be dealt with during the processing of the SRTM dataset. This article presents some facts about the void data, which were discovered processing the SRTM data.

Global Surface Model using Space-shuttle Radar Topographic Mission Dataset and Global Indexing Grid

Jan KOLÁŘ, Erik KJEMS, Lars BODUM, Esben Munk SØRENSEN, Denmark

1. INTRODUCTION

Modeling of the topographic surface has applications in various fields of civil engineering, urban planning, tele-communication and in a number of other domains of human activities. In turn methods and software components, which provide topographic models, gained a corresponding level of attention in research and development. However, with the expansion of three-dimensional (3D) geographic applications the representation of the surface has become more essential. Mainly because in these applications the role of the topographic surface is of similar importance as the role of the mapping plane in 2D cartography---it is a defining surface, which relates to all other geographic elements in the application. Terrain visualization turned into a distinguished subject in the field of computer graphics and with advances addressing multi-resolution representations computer graphics has become a major contributor to emerging geographic media of virtual globes.

From the perspective of this new 3D geographic technology, which is also the perspective taken in this article, the design of the technology for terrain modeling may be based on various methods that have been developed during the last decade. For instance a planetary representation of terrain using bin-tree cells has been introduced in (Cignoni et al., 2003). Each cell was filled with a triangular irregular network (TIN) consisting of a few hundred triangles. This method is known as batched dynamic adaptive meshes. Another framework focusing primarily on multi-resolution visualization is introduced in (Lindstrom and Pascucci 2002), which also provides an excellent overview of the achievements till that time. In the work of (De-Floriani et al., 2000) you will find a description of an extensible multi-resolution TIN system suitable for terrain analysis and processing rather than for a mere visualization. All these works support so called out-of-core mode of operation for maximum visualization performance. On the other hand another standpoint is taken in works of (Rabinovitch and Gotsman, 1997) or (Widmark, 2001) . They provide generally applicable remedy to the inexistent topological correlation on edges between different terrain resolutions by deploying a procedural mode of operation, which generates the geometry for rendering at run-time rather than out-of-core. Unlike all the works mentioned above (Lukatela, 2000) introduces a TIN-based seamless global terrain model, which avoids the use of spatial data structures defined on top of a 2D plane, i.e. his method uses only unprojected coordinates at the level of data representation.

Clearly there are myriad of aspects regarding the representation of the topographic surface for 3D geographic applications. But is there any uniform and generally applicable solution? Not today. Nevertheless, in order to contribute the efforts aiming at a more positive answer to that question, this work presents results based on yet another method described in (Kolar, 2006), which allows for instant 3D visualization over the Internet.

The article has following structure. First the method overview is given in the following section. Then a publicly available terrain data product SRTM, which has been used in the experiment, is shortly introduced in Section 3.1. An experiment description as well as some statistics about the results especially about the occurrence of void data is presented in Section 3.2. The conclusion is presented in Section 4 along with the main advantages as well as disadvantages revealed by the results and some research directions are given in the last section.

2. GIG INDEXING AND TERRAIN REPRESENTATION

The experiment presented in this article is based on implementation of global indexing grid (GIG) and a method for analysis and representation of the topographic surface. The first method allows for spatial indexing based on subdivision of the spherical surface. The second method addresses analysis and representation of the topographic surface data, which can (but does not necessarily have to) use the indexing mechanism. Overview for both concepts is presented in the following sections.

2.1 Global Indexing Grid

The global indexing grid (GIG) is a discrete global grid system first introduced in (Kolar, 2004) and further elaborated in (Kolar, 2006). GIG provides direct subdivision of the three-dimensional space, as an opposite to methods which deploy projections to Euclidian two-dimensional domain in order to subdivide the space. GIG is based on Dirichlet tessellation of the sphere, which was already suggested as a method for geographic indexing in (Lukatela, 1987). Unlike the previous works based on Dirichlet tessellation, GIG allows for a multi-resolution tessellation.

The elementary indexing unit in GIG is a cell, which has a shape of a cone. A cell's vertex is at the origin of the coordinate system and its boundaries go to infinity as depicted in Figure 1.a. Each cell is associated with a centriod, which are distributed on the unit sphere and have defining role for the cells. Intersection of the cell boundaries and the unit sphere marks off the Dirichlet tessellation on the sphere. There are GIG levels in this method, which provides means to multi-resolution tessellation. Each GIG level is defined by a single parameter---the division coefficient. In Figure 1.b are shown two GIG levels. The coarse one has division coefficient ten and the higher resolution has division coefficient set to three hundred. The division coefficient corresponds to the number of cells between poles along the zero-meridian.

The resulting database created during the experiment exploits GIG indexing for fast access to any part in the global model of the topographic surface. Among the main advantages of GIG is

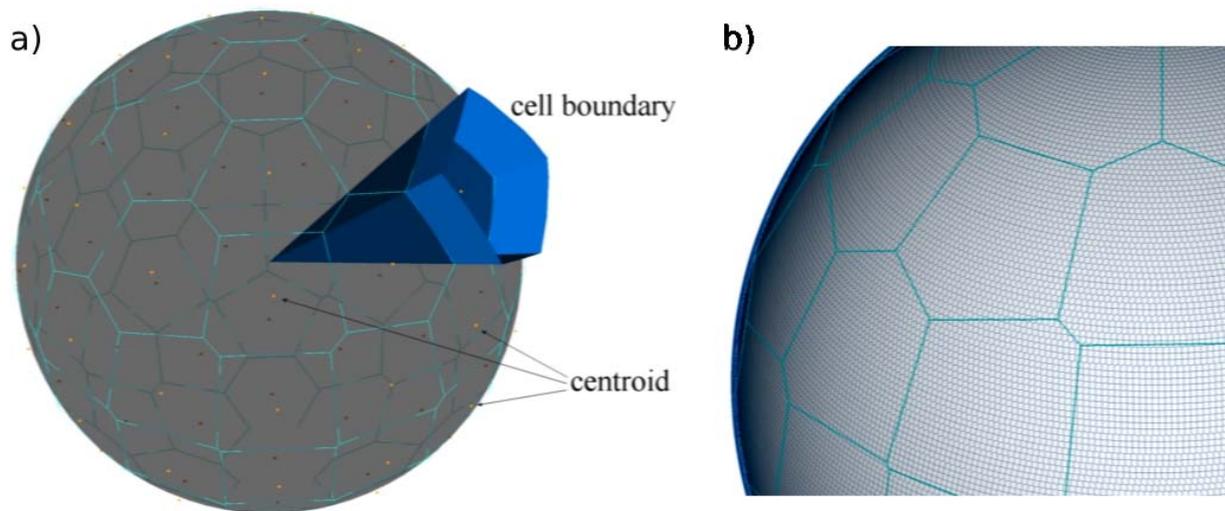


Figure 1. Schema of the Global Indexing Grid.

its support for the proximity spatial queries, which are common in many geographic applications and are also suitable for visualization. Dirichlet tessellation subdivides the space based on proximity to the centroids. Therefore GIG is efficient in retrieving data at a given location and its surroundings. However an unpleasant feature of GIG is that higher resolution levels cannot be obtained by recursive subdivision from the previous GIG level.

2.2 Representation of Topographic Surface

2.2.1 Data representation

The data representation for the topographic surface model is based on clusters of elevation points. This makes the representation simple and robust. The main attribute to the simplicity is the fact that data about surface topology are avoided entirely. This is convenient for multi-resolution models because there is no topological correlation between topographic surfaces at two or more resolutions, which often becomes problematic when parts of the same model, but at different resolutions, need to be stitched together. Additionally avoiding topological data makes the representation flexible for editing and is suitable for applications with storage distributed over many hosts.

The clusters of the stored elevation points correspond to the GIG cells introduced in Section 2.1.

The data representation is associated with Delaunay triangulation, which is used to reconstruct the surface topology at runtime. Two-dimensional Delaunay triangulation is performed against the horizontal plane at the given location. Formal description of the data representation is in (Kolar, 2006).

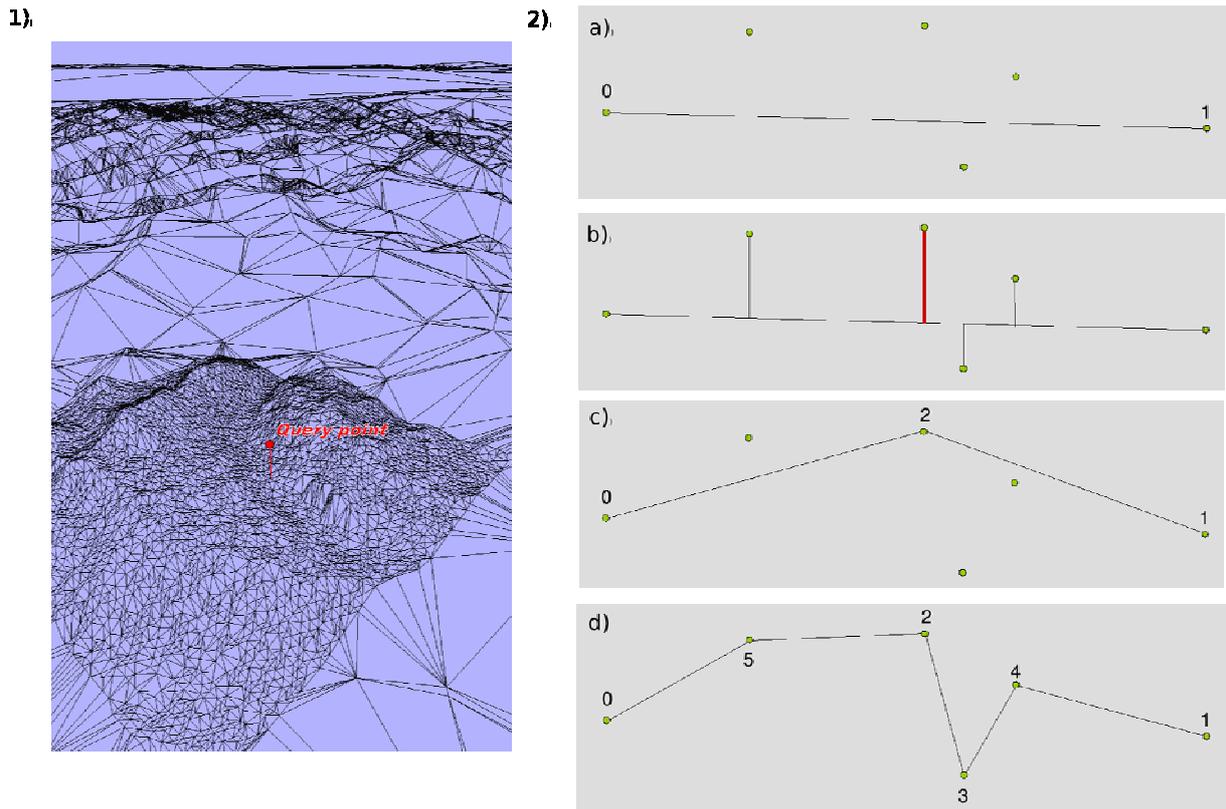


Figure 2. Processing of elevation data for multi-resolution surface representation.

2.2.2 Analysis of terrain data

In order to construct a multi-resolution model, a technique has been devised that could be applied to various terrain datasets and turn them into the representation described in the previous section. The main goal of the processing was to enable selection on elevation points. Elevation points for the low resolution part of the model will be selected for clusters in the coarser GIG levels and high resolution data in the GIG levels with smaller cells.

The analytical processing keeps all necessary elevation points from the input and sorts them by certain importance measure. Delaunay triangulation has also been exploited in the evaluation process. The process begins with an initial approximation by two triangles covering the same area as the input data and iteratively adds points to the triangulation. At each iteration the point (or set of points) with the highest importance is inserted into the triangulation. The process is demonstrated using a terrain profile as depicted in Figure 2.2. The most important point in the initial approximation (see Figure 2.2.a) is the one with the highest vertical distance from the approximation (highlighted in Figure 2.2.b). Hence the vertical distance is the importance measure for sorting the points.

The process of refinement continues until a desired vertical error threshold V is reached or all points are exhausted (see Figure 2.2.d). Similar method is described in (Garland and Heckbert, 1995).

Insertions of points are irrevocable; therefore the resulting list of points in the triangulation is sorted by the importance measure. This property is convenient for the construction of a multi-resolution model, because the whole process of refinement needs to be passed just once for the requirements on the best resolution. Arbitrary number of levels of resolution can be created by mere splitting of the list producing a TIN approximation. Example with three levels of resolution, as they were used in the experiment, is shown in Figure 2.1.

3. EXPERIMENT WITH SRTM

The method described in the previous section was used for the creation of near global coverage model of a topographic surface. The implementation of the method was made for the GRIFINOR platform (www.grifinor.net). This section provides the results from an experimental processing and construction of the model. Data from the Shuttle Radar Topography Mission (SRTM) were used for two main reasons: 1) SRTM data are publicly available and thus it is easy for others to relate these results to later results; 2) SRTM has a near global coverage.

3.1 SRTM Dataset

The Shuttle Radar Topography Mission (Farr and Kobrick, 2000) was a joint venture of NASA's Jet Propulsion Laboratory (JPL), National Imaging & Mapping Agency (NIMA, which has since then been renamed as the National Geospatial-Intelligence Agency (NGA)), the German Space Agency (DLR) and Italian Space Agency (ASI). The mission collected 12 terabytes of data over 80% of the earth's landmass in February 2000. Approximately 95% of the targeted landmass was imaged twice and 50% three times. The SRTM DEM product derived from interferometric analysis of the C band signal and was processed by NASA. The X band signal and its derived products were processed by DLR and ASI. The SRTM DEMs are now being distributed by several agencies both public and private in various formats, however their availability changes rapidly. For the experiment described further in the text were used DEMs distributed by the USGS at <ftp://e0srp01u.ecs.nasa.gov/srtm>. This product is based on the C signal and is provided in three levels of resolution, namely thirty arc-second (SRTM30), three arc-second (SRTM3) and one arc-second (SRTM1). SRTM3 provides horizontal resolution that corresponds to 90m between samples along the equator. SRTM3 is derived from the SRTM1 data by the averaging of 3x3 elevations samples. Height values for the SRTM dataset are related to the EGM96 geoid and the grid is referenced by geographic latitude and longitude on the WGS84 ellipsoid. This data are described as being of "research grade". Due to the acquisition technique of the interferometric synthetic aperture radar (InSAR) the SRTM data include voids and spurious height values. Furthermore water bodies often have a rough appearance. Therefore this dataset cannot be used for critical applications in practice without correction processing and quality checking.

3.2 Experiment with GIG and SRTM

In the experiment the SRTM3 product was used. This dataset contains the highest available resolution at near global coverage. The first version of SRTM3 was completed in July 2004. The dataset used in the experiment was the second version available since July 2005. This version contains improvements such as smoothing large water areas or "flattening" all data from the sea surface to have the value zero. Each file covers one by one degree in geographic coordinates and contains raster data with 1201x1201 samples. This dataset was released on a continent by continent basis. The tiles are organized on the DTED1 model with the tile name referring to the longitude and latitude of the south west corner of the tile.

3.2.1 Setup and basic statistics

The overall amount of 14 572 files counting all available SRTM3 data files were processed on two computers more or less simultaneously. The first computer C1 had two AMD Opteron processors working at 2GHz, the second computer C2 had two dual-core IBM PowerPC 970MP (G5) processors at 1.8 GHz. Each of the computers had one GB of RAM per processor and Java 5 virtual machine. Computer C1 processed 9152 files and C2 processed 5420 files. For each file/tile six measures were recorded;

- file name (coordinates of the tile)
- number of void pixels in the tile
- number of boundary pixels around the void area
- number of elevation points after surface analysis
- time for storage in the indexed database in seconds
- overall time for file processing in seconds

Thanks to the method (see Section 2) the void pixels just can be dropped when detected during the processing. That is far more elegant as well as more correct than "inventing" the missing values, which must be done for the surface representations using raster data structures. Despite the fact that the GIG method is immune to void values they obviously remain harming the quality of the resulting model as the data are simply missing. The average amount of void samples in each SRTM file was 24539.45, which means that 1.7% of the SRTM data are voids. Thirteen files turned out to contain only void data. On the other hand 4187, which stands for 28.7% of all files, are "clean" tiles that are free of any voids. The ratio between the number of void pixels in the tile and the number of boundary pixels around the void areas suggests that voids tend to form small clusters rather than large coherent "holes".

For the entire experiment the vertical error threshold (see Section 2.2.2) was set to $V=1$, which corresponds to one meter using the units from the original raster model. With this parameter the resulting global model of the topographic surface consists of 8 501 099 736 (approximately eight and half billion) elevation points. This means that the number of the used elevation points decreased to 40.4% compared to the original SRTM files, or in other terms the result needs 2.47 times fewer points.

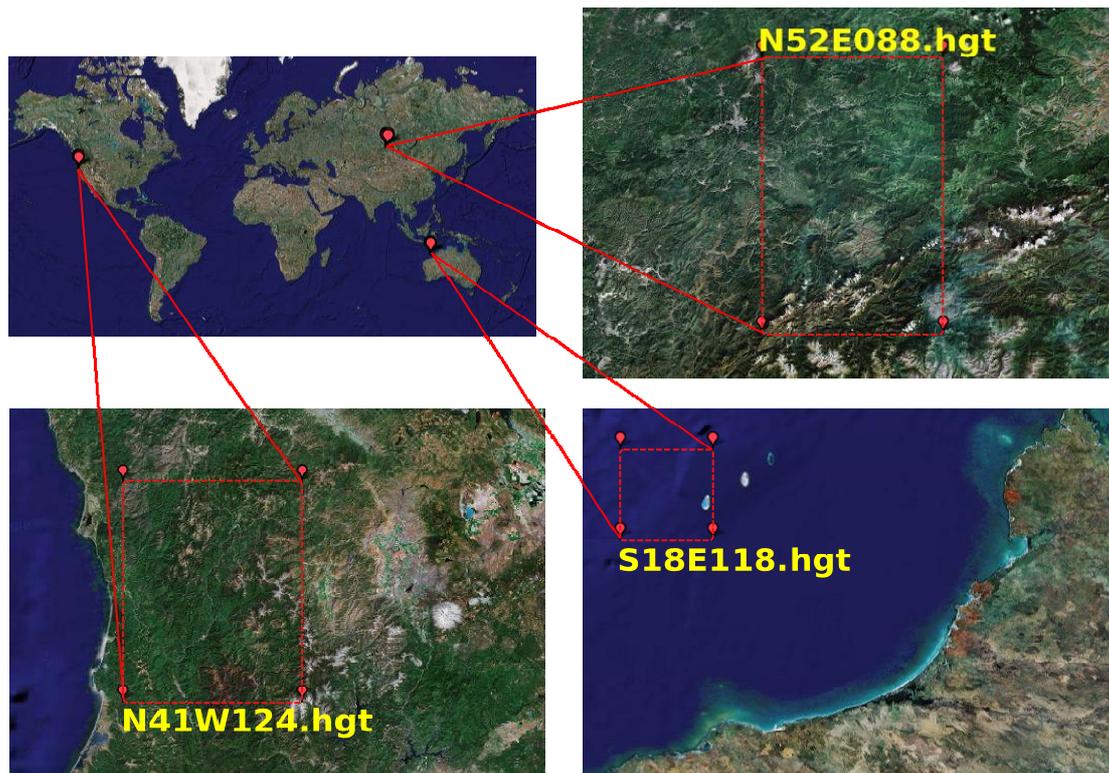


Figure 3. SRTM tiles with the "roughest" and "smoothest" topographic surface.
(data by Google)

3.2.2 Influence of surface geometry

The measure about the number of elevation points after the surface analysis delivers values that are proportional to the "roughness" of the topographic surface, i.e. how frequently elevation changes non-linearly between the samples from the original raster DEM. There was no correlation found between the resulting number of points in a tile and the location of the tile. There were three tiles in which all the original points were eliminated. These tiles are namely:

- Australia/S18E118.hgt
- Africa/N16E055.hgt
- Eurasia/N49E155.hgt

The topographic surface in these areas is represented in the resulting model by only four points at the corners of the tiles. For example, the area covered by S18E118.hgt is depicted in Figure 3. Apparently this surface might indeed be smooth as most of the area is covering the surface of the ocean with an atoll-island, which most probably is closely aligned with the ocean (according to our setup up to one meter).

On the other hand the “roughest” topographic surface was during the experiment detected in 'North_America/N41W124.hgt' file. This tile was represented in the resulting model by 1 002 071 points, which is 69.5% compared to the 1201x1201 sized samples. The “roughest” tile free of any voids was Eurasia/N52E088.hgt with 926 228 elevation points (64.2%). The locations of these tiles are also depicted in Figure 3.

According to our knowledge about the origin of the void values, the correlation between the number of resulting points and the number of voids should indirectly address the relationship between the “roughness” of the topographic surface and (in)capability of the InSAR technology to scan the surface. Two diagrams were generated regarding this matter, both showing the number of void data in a tile (the tile is represented by a point in the diagram) in relation to the number of resulting elevation points in that same tile. The first diagram, which is shown in Figure 4., uses y-axis scaled in correspondence to the average number of voids in the entire SRTM3 dataset (see above for the average). It can be observed that with increasing number of elevation points in a tile there is an increasing amount of voids. This shows that roughness of the surface causes more voids.

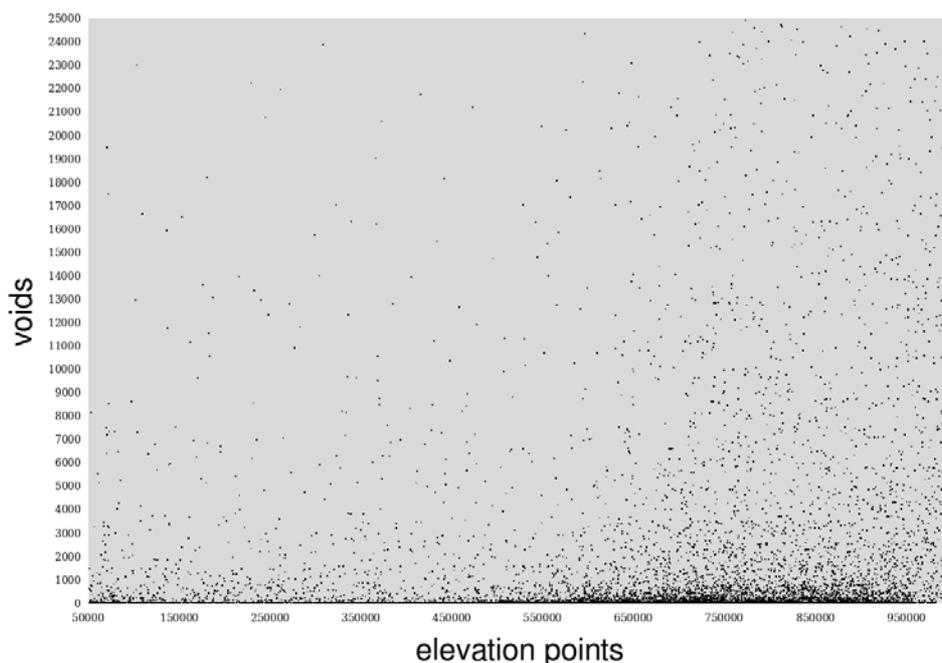


Figure 4. Dependency of void samples on number of resulting points in a tile (scaled).

The second diagram shows the same data, but on an interval (0; 1 442 401). That is the largest theoretically possible range for both measures since the maximum number of points in each input file is $1201 \times 1201 = 1\,442\,401$. Also note that voids and the resulting number of points are mutually exclusive measures in this experiment depicted by the diagonal (all possible combinations are below the diagonal). This diagram also reveals another correlation showing a cluster of tiles counting about 900 000 voids (marked with a ring in Figure 5). This cluster however must have a different explanation compared to the first diagram.

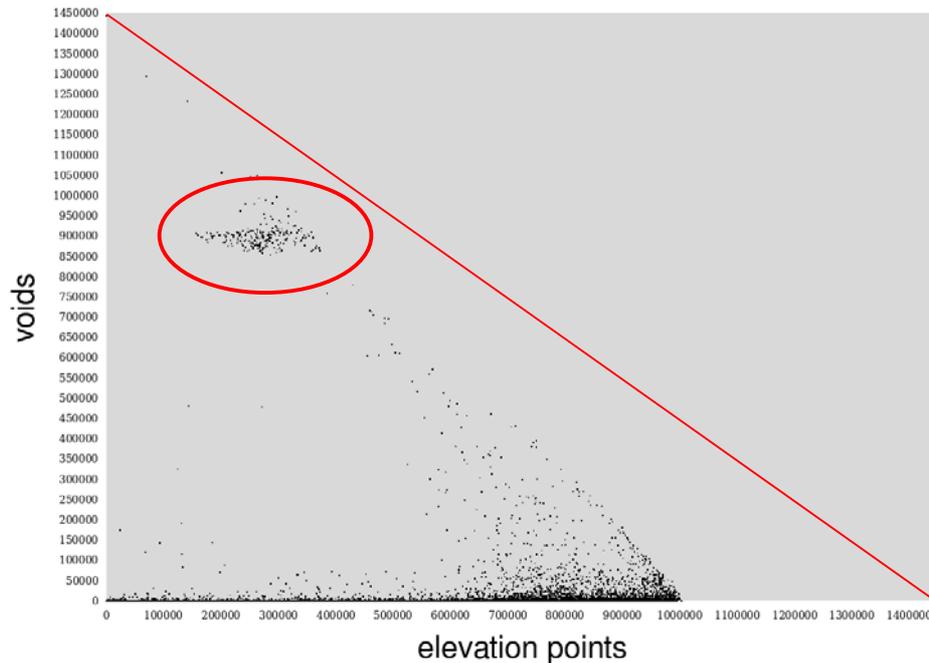


Figure 5. Dependency of void samples on number of resulting points in a tile.

3.2.3 Database construction

The resulting points have been stored in an indexed database as Java objects. The main purpose of this database is to provide the resulting model to the GRIFINOR platform; an Internet-based 3D geovisualization technology (www.grifinor.net). Each object in the database represents a single GIG cell (see Section 2) and contains only elevation points from that cell. The indexing was made at three levels of resolution using GIG levels 100, 1500 and 3000. The GIG levels were selected empirically from visualization of the data in 3D scene. Therefore selection of GIG levels was made without any theoretical ground. The storage of the resulting model was addressed by the absolute measure of time spend for the following processing tasks:

- Separation of points into corresponding GIG levels
- Conversion of normal height to ellipsoidal height
- Transformation to geocentric coordinates
- Indexing geocentric coordinates to a GIG cell
- Storage of the cell data.

The size of the resulting database is 99 GB of data, which is 2.5 times more compared to the original SRTM3 dataset, which has only 40 GB of data.

Just for curiosity the entire experiment was processed in 3.5 days. But this number cannot be used as a measure as it depends on a number of variables, e.g. the number of computers/processors being used, the processing power of each CPU, speed of disks etc. In the experiment, C2 for example only contributed by 47.2 hours of processing time during the experiment and only one of the processors was actually occupied all the time.

The Computer C1 performed an average of 27418.2 indexed points per second, while computer C2 only indexed 18033.4 points per second.

4. CONCLUSIONS

In this article an experimental processing of SRTM3 data for sake of constructing a near-global model of the topographic surface has been presented. The experiment exploits GIG method for spatial indexing and topology-less data representation that is associated with Delaunay triangulation. The database resulting from this experiment produces TIN model at three levels of resolution and delivers data for instant visualization in 3D over the Internet. This result has been published at www.grifinor.net and can be seen by broad public.

Although the used processing method efficiently eliminates void data samples that are present in the SRTM, the experiment reveals some statistical information associated with voids in SRTM3 dataset as a whole. It has been found that in average 1.7% of the SRTM data are missing data and that only 28.7% of the distribution files are complete---without any voids. It has been found that rugged topography means more missing data in SRTM. This result was expected and supports theory about InSAR technology used for acquisition of SRTM. There has also been found a significant amount of SRTM3 tiles in which about 62% of the data are voids. This statistical result is left without any reasonable explanation.

The resulting database accounts 99GB of data and is 2.5 times bigger than the original SRTM3 dataset. This is considered a satisfactory result due to the properties that were gained in the new topographic database. In summary these properties account: 1) a uniform three-dimensional coordinate system; 2) a single continuous model; 3) low level data representation suitable for analytical processing; 4) a multi-resolution model; 5) ability to add or update a part of the model; 6) the ability to instantly visualize the surface from the database and; 7) the ability to manage the model on multiple servers in a decentralized manner. Another advantage of the experiment is its general design. It is applicable to most DEM or DTM data acquired today. The input can be raster data, TIN data or mere point cloud with small changes. It also partially addresses the obstacle of erroneous data such as the voids in SRTM. Similar problems are common for an output from contemporary scanning and ranging technologies.

The main disadvantage might be considered performance, which cannot match solutions based on raster data or on representations that store surface topology. Data representation for break lines or enforced edges are missing in the described experiment, but it is possible to add their support in the future. Investigation of results using a different error threshold V is also a suggestion for the future experiments. But there are other hints for improvements to the presented experiment organization. For example it would be interesting to trace relative complexity of the tasks in the database construction described in Section 3.2.3. On the other

TS 5H – Data Modeling and Algorithms

11/13

Jan Kolar, Erik Kjems, Lars Bodum and Esben M. Sorensen

Global Surface Model using Space Shuttle Radar Topographic Mission Dataset and Global Indexing Grid

Strategic Integration of Surveying Services

FIG Working Week 2007

Hong Kong SAR, China, 13-17 May 2007

hand the information about boundary of the void areas was not very useful yet relatively computationally demanding.

REFERENCES

- Cignoni, P., Ganovelli, F., Gobbetti, E., Marton, F., Ponchio, F. and Scopigno, R. (2003). Planet-sized batched dynamic adaptive meshes (p-bdam), Proceedings IEEE Visualization, IEEE Computer Society Press, Seattle, WA, USA, pp. 147–155.
- Farr, T.G. and M. Kobrick, 2000, Shuttle Radar Topography Mission produces a wealth of data, Amer. Geophys. Union Eos, v. 81, p. 583-585.
- De-Floriani, L., Magillo, P. and Puppo, E. (2000). Variant: A system for terrain modelling at variable resolution, Geoinformatica 4(3): 287–315.
- Garland, M. and Heckbert, P. (1995). Fast polygonal approximation of terrain and height fields, Technical report, Carnegie Mellon University. CMU-CS-95-181.
- Kolar, J. (2004). Global indexing of 3d vector geographic features, International Society for Photogrammetry and Remote Sensing 20th Congress, Istanbul, Turkey, pp. 669–672.
- Kolar, J. (2006). Global indexing of topographic surface for 3-d visualization and analysis, Ph.D. thesis, Aalborg University.
- Lindstrom, P. and Pascucci, V. (2002). Terrain simplification simplified: A general framework for view-dependent out-of-core visualization, IEEE Transactions on Visualization and Computer Graphics 8: 239–254.
- Lukatela, H. (2000). A seamless global terrain model in the hipparchus system, International Conference On Discrete Global Grids, Santa Barbara, California, USA.
- Rabinovich, B. and Gotsman, C. (1997). Visualization of large terrains in resource-limited computing environments, IEEE Visualization '97 pp. 95–102.
- Widmark, M. and Blueberry3D (2001). Procedural geometry for real-time terrain visualisation in blueberry3d, Digital Earth Moving: First International Symposium, Manno, Switzerland.

BIOGRAPHICAL NOTES

Jan Kolar, Ph.D., M.Sc., Ing., is an assistant professor in the Centre for 3D GeoInformation at Aalborg University. Jan received his master's degree in geodesy and cartography at Czech Technical University in 2001, master's degree in computer science at Aalborg University in 2002 and doctoral degree in planning and development at Aalborg University in 2006.

TS 5H – Data Modeling and Algorithms
12/13

Jan Kolar, Erik Kjems, Lars Bodum and Esben M. Sorensen
Global Surface Model using Space Shuttle Radar Topographic Mission Dataset and Global Indexing Grid

Strategic Integration of Surveying Services
FIG Working Week 2007
Hong Kong SAR, China, 13-17 May 2007

Erik Kjems, Ph.D., M.Sc., B.Com., is an associate professor at the Department of Development and Planning, Aalborg University, Denmark. He is a member of the Centre for 3D GeoInformation.

Lars Bodum, Ph.D., M.Sc., is an associated professor and director for Centre for 3D GeoInformation at Aalborg University. He graduated as chartered surveyor (M.Sc.) in 1990 and received his Ph.D. in 2000 also at Aalborg University. Currently he is a faculty member of the Department of Development and Planning, Geoinformatics where his research interests concentrate on Spatial Multimedia and Virtual Reality, 3D Geoinformation, Geographic Information Technology in Spatial Planning, Digital Planning Documents and GIS on the Internet.

Professor **Esben Munk Sørensen**, is professor in Land and GeoInformation Management at Aalborg University since 1998. He published more than 100 articles on GIS, Rural Development and Land Consolidation, Land Management and Spatial Planning, and on virtual learning environment. He was invited speaker with paper on FIG Congresses and Workingweeks several times since 1998.

CONTACTS

Jan Kolář, Erik Kjems and Lars Bodum
Centre for 3D GeoInformation
Aalborg University, Fibigerstraede 11
9220 Aalborg
DENMARK
Tel. +45 9635 9797
Email: {kolda,kjems,lbo}@3dgi.dk
Web site: <http://www.3dgi.dk>

Esben Munk Sørensen
Aalborg University, Fibigerstraede 11
9220 Aalborg
DENMARK
Tel. +45 9635 8347
Email: ems@land.aau.dk
Web site: <http://www.land.aau.dk/~ems>