

The Optimum Dataset method – examples of the application

Wioleta Błaszczak-Bąk¹, Anna Sobieraj-Żłobińska^{2,*}, and Beata Wieczorek³

¹University of Warmia and Mazury in Olsztyn, Faculty of Geodesy, Geospatial and Civil Engineering, Institute of Geodesy, Oczapowski St. 1, Olsztyn, Poland

²Gdańsk University of Technology, Faculty of Civil and Environmental Engineering, Department of Geodesy, Narutowicza St. 11/12, Gdańsk, Poland

³University of Warmia and Mazury in Olsztyn, Faculty of Geodesy, Geospatial and Civil Engineering, Department of Land Surveying and Geomatics, Heweliusza St. 12, Olsztyn, Poland

Abstract. Data reduction is a procedure to decrease the dataset in order to make their analysis more effective and easier. Reduction of the dataset is an issue that requires proper planning, so after reduction it meets all the user's expectations. Evidently, it is better if the result is an optimal solution in terms of adopted criteria. Within reduction methods, which provide the optimal solution there is the Optimum Dataset method (OptD) proposed by Błaszczak-Bąk (2016). The paper presents the application of this method for different datasets from LiDAR and the possibility of using the method for various purposes of the study. The following reduced datasets were presented: (a) measurement of Sielska street in Olsztyn (Airborne Laser Scanning data – ALS data), (b) measurement of the bas-relief that is on the building in Gdańsk (Terrestrial Laser Scanning data – TLS data), (c) dataset from Biebrza river measurement (TLS data).

1 Introduction

Rapidly developing measurement technologies such as LiDAR (Light Detection And Ranging), MBES (Multi Beam Echo Sounder), bring large amounts of data. With such datasets, digital models of measured objects can be generated, for example: Digital Terrain Model or the Digital Elevation Model. However processing of such amount of data, especially if there are real-time changes occurring, is virtually impossible or very difficult [14]. Constantly, there are ongoing works on improving the processing of large datasets. Such datasets can be used to build the Spatial Information System (SIS), as well as being the source of many studies. One of the projects that aims to provide an advanced use of spatial data is the Center for Spatial Analysis of Public Administration (Centrum Analiz Przestrzennych Administracji Publicznej - CAPAP). Within the project tasks and tools related to spatial data processing are planned. The project is planned to include execution of tasks related to tools enabling the use of spatial data. One of such tool is an analytical platform that enables advanced spatial analysis, including 3D data analysis, as well as interpretation and visualization of the results in textual and graphical form.

In this and many other cases [e.g.: 1, 16] it is necessary to reduce the the measurement set. Decrease of the data can be conducted by means of reduction or generation. Generation is decreasing the dataset by creating a grid. In this method we have new points instead of points with the original coordinates [2, 11]. While reduction decreases the dataset by removing some points according to the given algorithm, the remaining

points are original points from the measurement [4, 5, 6, 9]. For people using data in the form of point clouds, it is better and easier to use real data. Therefore, reduction is a better option.

Reduction of the dataset is a problem that needs to be properly planned so that the dataset after the reduction meets all the user's expectations. Evidently, it is best if the result is the optimal solution for the adopted criteria. It can be achieved by using the Optimum Dataset (OptD) method [7, 8] - the optimum method of reduction. This paper presents various variants of this method's applications. The obtained reduced datasets were generated on the basis of data from: (a) measurement of the Sielska street in Olsztyn (Airborne Laser Scanning data – ALS data), (b) measurement of the bas-relief that is on the building in Gdańsk (Terrestrial Laser Scanning data – TLS data), (c) dataset from Biebrza river measurement (TLS data).

2 The Optimum Dataset method

The OptD method is designed to reduce large sets of data such as ALS data and TLS data. The method takes into account the different levels of reduction in the individual parts of the processing area. And as a result, there are more points in detailed part of scanned object. In the case of uncomplicated structures or areas, the number of points is much less. Only those points that are significant will remain, and the generated model will meet the predetermined parameters, for example, the accuracy of the obtained model. The method has been described in details and presented in [7, 8].

* Corresponding author: annsobic@pg.edu.pl

The most important step in the OptD method is the selection of a line generalization algorithm that creates points in the measurement strips with the coordinates (x_i, z_i) for vertical strips or with the coordinates (y_i, z_i) for horizontal strips. Research shows that it is enough to use the generalization algorithm in the chosen system. There is no need to create vertical and horizontal strips at the same time.

During the reduction, the position of the points relative to each other is considered. The selected algorithms presented for example in [10, 13, 18] determines the degree of reduction by applying the appropriate tolerance range.

3 Examples of applications

3.1. DTM generation

Digital Terrain Model DTM is one of the most popular product which can be used in many fields of economy and science. In order to created it a reliable data is necessary. One of the source for DTM generation is point clouds obtained during laser scanning measurements.

The study area for which airborne laser scanning measurement was conducted is a fragment of the national road No. 16, a Sielska street in Olsztyn, located in the Warmia-Mazury. Measurement made by Visimind Ltd enabled the acquisition of point clouds. A fragment of the original ALS dataset (144500 points), which was used as a study area of this research, is presented in Fig. 1a.

The selected fragment was filtered by using adaptive TIN model method implemented in own software. As a result of the filtration, there are two sets of data: a) the set of points showing the topography (108313 points), b) a set of points showing the details points (36187 points). Point cloud after filtration, which was used to generate the DTM is shown in Fig. 1b.

Point cloud after filtration comprising only ground points was optimized by OptD-single method. Application of the OptD-single method selected the optimum solution, which is presented in Fig. 1c.

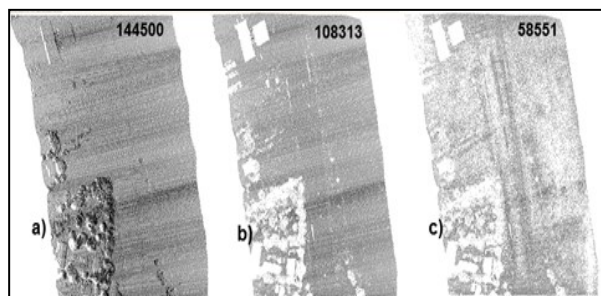


Fig. 1. Fragment of ALS point cloud original, b) after filtration, c) after the OptD-single method was applied (source: Błaszczak-Bąk 2016).

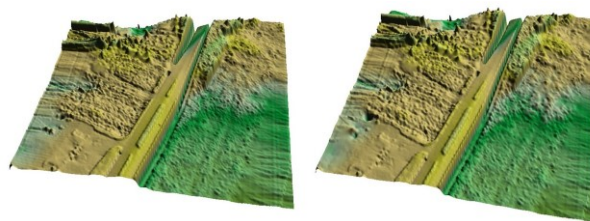


Fig. 2. DTMs generated on the basis of: the original set consisting of 108313 points (left), the set decreased by means of the OptD-single method consisting of 58551 points (right) (source: Błaszczak-Bąk 2016).

Standard deviation SD and coefficient of determination D (denoted also as R^2) were calculated for the generated DTMs. Coefficient of determination D is the measure of model adjustment (the closer to 1, the better the match of the model to another). The results are presented in Table 1.

Table 1. Parameters of generated DTMs.

| Parameters | DTM based on original data | DTM based on data decreased by OptD-single method |
|------------|----------------------------|---|
| SD [m] | 0.071 | 0.083 |
| D | - | 0.980 |

The coefficient of determination indicated good adjustment of DTM generated on the basis of reduced dataset to DTM based on original dataset. The SD is higher for DTM for OptD-single method only about 0.01m.

3.2 3D modelling

LiDAR technology makes it possible to obtain point clouds with a high density, resulted from the scanning resolution set by the operator. Issues related to the processing of LiDAR data can be found, among others, in [12], where automatic estimation of agricultural tree geometric parameters and its accuracy are presented, in [14] who described the processing of 3D modelling triangle meshing. In paper [3] the authors discussed the problem of 3D modelling of terrestrial objects. In all of these papers the whole point cloud was used and the density of points was relatively the same throughout the scanned object. Very often such amount of data and uniform density is not needed, especially when the main focus is on the features of relatively flat object (e.g.: bas-relief) [2]. For this purpose, it is worthwhile to use the OptD method [7]. It allows to reduce the obtained dataset without losing the characteristic points and in result the density of points is various throughout the scanned area. Application of the OptD method, especially for objects with complex structure, can help to reduce the number of points without losing the information necessary for proper modelling.

As a research facility, the building located within the University of Gdansk was used. On the front elevation (north wall) there is a bas-relief. It shows the image of a

dragon and a man coal-stoker with a shovel. The fragment of obtained point cloud including the bas-relief consists of 753583 points. The whole bas-relief is shown in Fig. 3.

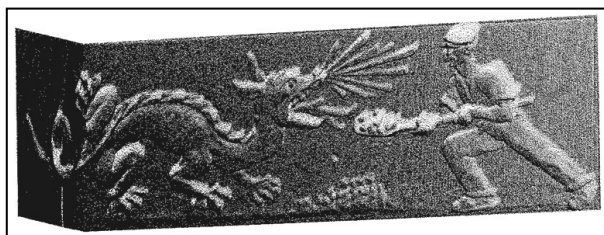


Fig. 3. Point cloud with bas-relief (source: own study in the CloudCompare v.2.7.0).

TLS point cloud with bas-relief was processed by means of the OptD-single method. After optimization 488715 points remained. It is about 64 % of the points of the original point cloud.

Fig. 4 presents a bas-relief with lower density of points, however a dragon and a coal-stoker with a shovel are still very well visible. The characteristics details are still preserved.

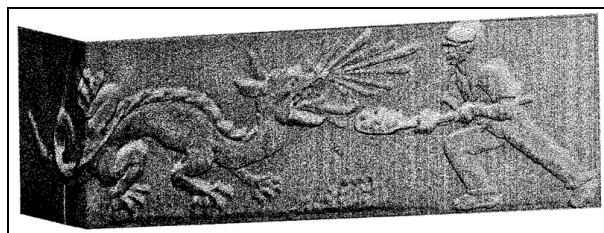


Fig. 4. Point cloud after reduction (source: own study in the CloudCompare v.2.7.0).

Decreased dataset was used to build 3D model mesh. A mesh is a set of vertices, edges and faces that defines the shape of a polyhedral object in 3D computer graphics. The faces usually consist of triangles (triangle mesh), quadrilaterals, or other simple convex polygons, since this simplifies rendering, but may also be composed of more general concave polygons, or polygons with holes [17].

On the basis of original TLS dataset and dataset after application of the OptD method the 3D model were generated. In Fig. 5a and 5b original and reverse views of the bas-relief are presented. The reverse view was chosen to use because the details of bas-relief are much better visible. Models were generated in CloudCompare v.2.7.0 as mesh.

Generated 3D models are almost identical. There are no significant differences. The advantage of 3D model generated on the basis of optimized dataset is the decreased number of points in input set, what in result accelerated the modeling. In the case of decreasing the set about 35.1 %, the time needed for modeling decreased about 33.3 % (360 sec for whole set, 240 sec for optimized set).

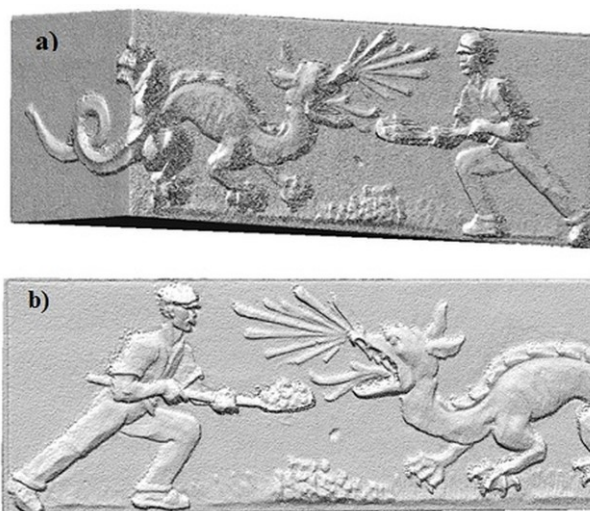


Fig. 5. Bas-relief 3D model generated on the basis of optimized dataset a) original view, b) reverse view (source: own study in the CloudCompare v.2.7.0).

In order to compare the obtained results and indicated the differences, the contour models based on GRID were also generated for fragment of the bas-relief (coal-stoker figure). They are presented in Fig. 6a and Fig. 6b.

For generated models OriDset (original dataset) and OptDset (optimized dataset) following assessment parameters were calculated:

- mean value of height difference (Δh_{mean}),
- coefficient of determination (D),
- standard deviation (SD).

Results are presented in Table 2.

Table 2. Comparison of generated GRIDs.

| Parameters: | Bas-relief GRID | |
|------------------------------|-----------------|---------|
| | OriDset | OptDset |
| Δh_{mean} [m] | -0.001 | |
| SD [m] | 0.234 | 0.245 |
| D | - | 0.99 |

Comparison of mentioned parameters shows that the coefficient of determination is very close to the value 1 for the OptDset. Thus, GRID generated from the reduced dataset is identical with GRID generated on the basis of the whole dataset representing the bas-relief, what means that interpolated heights in corresponding vertices are almost the same. Standard deviation SD calculated for OriDset and OptDset differs only about 0.001m, so it is almost negligible difference.

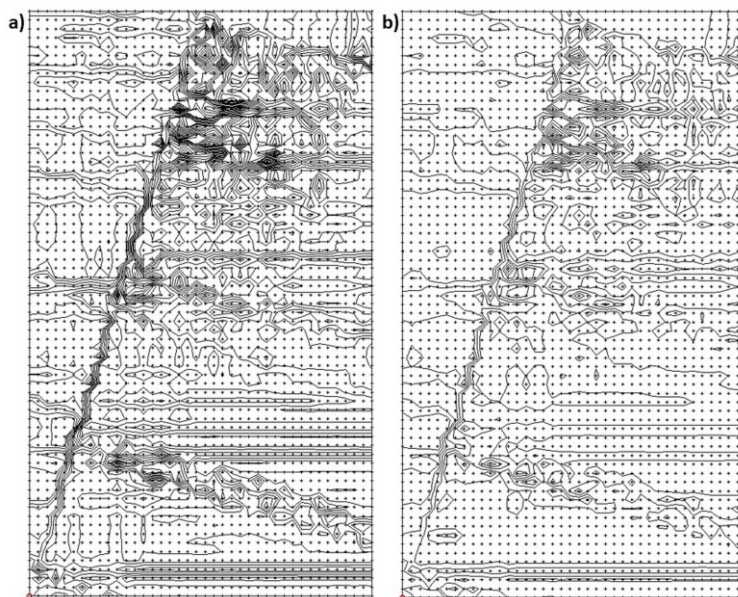


Fig. 6. The contour models for: a) original set, b) optimized set (source: own in the Surfer v.8).

3.3 Point cloud visualization

Point clouds and its visualization are good sources of information in case if there are changes within measured objects/surfaces. One of the example is meandering river area, where flow of inland waters, soil erosion and direct or indirect human activities are causing deformation of the land. Usefulness of spatial visualization of terrain in order to analyze such changes is determined by spatial data and methods of processing and modeling. Both, data and these methods, should allow to obtain a model with fixed or the appropriate geometric accuracy, in particular the altitude. The data and methods also should enables the appropriate mapping surface detail of the terrain model, wherein the amount of data need to be optimal. The large amount of data does not mean that visualization will be better or more accurate. Therefore, it was decided to apply the OptD method on the data acquired during measurement of one of meandering section of Biebrza river near village Goniądz conducted in May 2012. The C10 Leica Geosystem scanner was used.

In Fig. 7 there is a fragment of original TLS point cloud (from one station).

Characteristics of original and decreased sets are presented in Table 3.

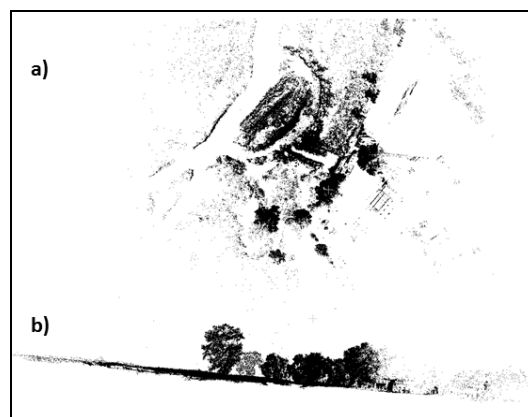


Fig. 7. The fragment of original TLS point cloud from one station for Biebrza river: a) top view, b) side view (source: own study in CloudCompare v.2.7.0).

Table 3. Characteristics of original and decreased sets.

| Characteristics | Original set | Decreased set |
|--------------------|--------------|---------------|
| Number of points | 1022857 | 451915 |
| Length [m] | 396.531 | 396.531 |
| Width [m] | 399.279 | 399.279 |
| Max height [m] | 76.714 | 76.714 |
| Min height [m] | -6.171 | -6.171 |
| H standard dev [m] | 4.036 | 5.521 |
| Mean height [m] | -0.826 | 0.550 |
| SD [m] | 4.036 | 5.521 |

* Corresponding author: annsobie@pg.edu.pl

The set was processed by means of the OptD-single method. There were 570942 points removed, what is 56 % of the original set. Parameters like length, width of processed area are the same, as well as Max height and Min height. It proves that the algorithm of the OptD method preserved the extreme values.

Decreased dataset is presented in Fig. 8.

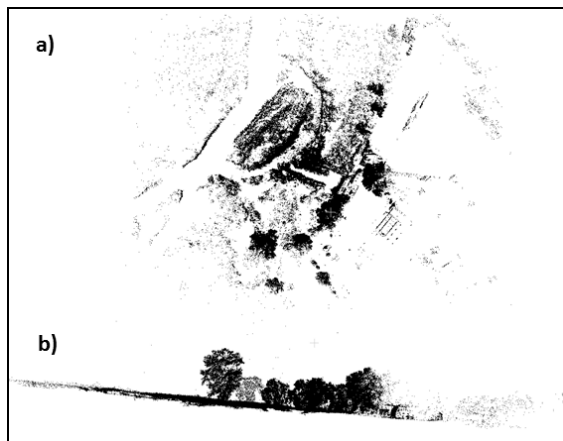


Fig. 8. The fragment of TLS point cloud from one station for Biebrza river after OptD method application: a) top view, b) side view (source: own study in CloudCompare v.2.7.0).

At the first sight, there is no significant differences. They are visible in zoomed fragment (Fig. 9).

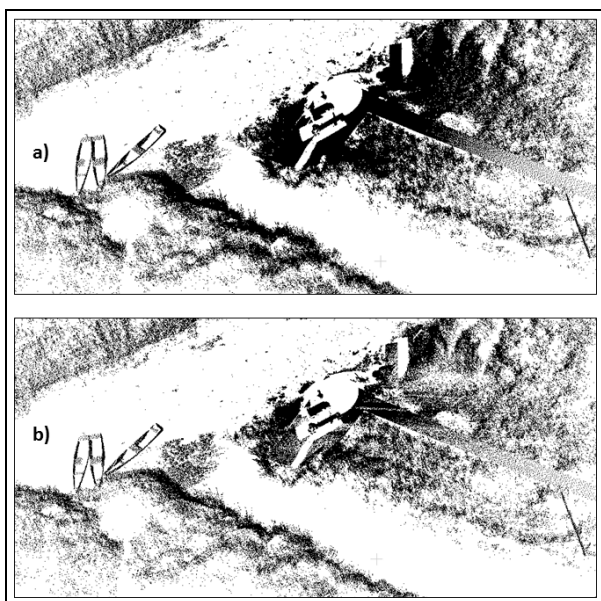


Fig. 9. The zoomed fragment of TLS point cloud for Biebrza river: a) original, b) after OptD method application (source: own study in CloudCompare v.2.7.0).

As it can be seen in Fig. 9 application of the OptD method resulted in more clear visualization, especially in those areas where there was very high density of measured points.

4 Conclusions

The OptD method works very well not only for ALS, but also for TLS. It is based on creation of strips (horizontal or vertical). Those strips are usually based on the measurement strips resulted from the way the object/area is scanned or could be determined by the object's shape, size, range etc. The strips can be analyzed in plane X_0Y (ALS and TLS, extensive object) or in Y_0Z or X_0Z (TLS, relatively small object). However, generalization algorithm works in each strip individually. The relative position of points within strip is analyzed in relation to Z coordinate (first case) or in relation to X or Y coordinate. In both variants the OptD method gives satisfying results.

On the basis of the presented studies showing the applications of the OptD method on TLS and ALS data, the following detailed conclusions were drawn:

- applying the OptD method for decreasing the datasets did not affected DTMs generated on the basis of these datasets – for ALS data coefficient of determination D is 0.98; for TLS data D is 0.99; differences in standard deviations SDs calculated for original and optimized datasets are 0.012m and 0.011m, respectively,
- difference in standard deviations SDs calculated for TLS data for extensive area is 1.485m. It resulted from the fact, that it is was relatively flat area where more points was removed, therefore mean heights for original and reduced datasets are different.
- comparison of meshes generated from original and optimized dataset for relatively small objects do not show significant differences, they are visible when grid is generated,
- visualizations of original and optimized datasets shows differences when it is zoomed.

And the general conclusions are:

- The OptD method reduces the ALS and TLS data.
- Reduction of the dataset in many cases improves its readability, and consequently its visualization.
- Reduced dataset can be used to build DTM.
- The OptD method can be used by architects to inventory objects with complex structure.
- The OptD method can be used in CAPAP project.

References

1. K. Bakula, *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **23**, 19 (2012)
2. B. Bauer-Marschallinger, D. Sabel, W. Wagner, *Computers & Geosciences* **72**, 84 (2014)
3. L. Barazzetti, F. Remondino, M. Scoioni, *Proceedings of SPIE Optics+Photonics* **7447** (2009)
4. W. Błaszczak, *Optimization of large measurement results sets for building data base of special information system*. Doctors thesis (University of Warmia and Mazury in Olsztyn, 2006)
5. W. Błaszczak-Bąk, A. Janowski, W. Kamiński, J. Rapiński, *Canadian Journal of Remote Sensing* **37** (6), 583 (2011)

6. W. Błaszczak-Bąk, A. Janowski, W. Kamiński, J. Rapiński, *Journal of Indian Society of Remote Sensing* **39**, 591 (2011b)
7. W. Błaszczak-Bąk, *Acta Geodynamica et Geomaterialia* **13/ 4(184)**, 381 (2016)
8. W. Błaszczak-Bąk, A. Sobieraj-Żłobińska, M. Kowalik, *Meas. Sci. Technol.* **28 (7)**, 075009 (2017)
9. Y. Chen, *High performance computing for massive LiDAR data processing with optimized GPU parallel programming* (2012)
10. D. H. Douglas, T. K. Peucker, *Canadian Cartographer* **10(2)**, 112 (1973)
11. D. Gościewski, *GIScience&Remote Sensing* **50(5)**, 515 (2013)
12. E. Hadaś, A. Borkowski, J. Estornell, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLI-B8**, 629 (2016)
13. G. F. Jenks, *Annals of the Association of American Geographers* **17 (1)**, 1 (1981)
14. T. Kersten, K. Mechelke, L. Maziull, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science* **XL-5/W4**, 1 (2015)
15. H. Liu, X. Yang, *Software reuse in the emerging cloud computing era*, chapter 9 (Information Science Reference, 2012)
16. M. Pomykoł, M. Poniewiera, *Numeryczne projektowanie w geodezji górniczej* (Wyd. Politechniki Śląskiej, Gliwice, 2009)
17. F. Remondino, F. Menna, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science* **XXXVII (B5)**, 199 (Beijing, China, 2008)
18. M. Visvalingam, J. D. Whyatt, *Line generalization by repeated elimination of point* (Cartographic Information Systems Research Group, University of Hull, 1992)