PROPOSED TECHNOLOGY OF LiDAR DATA PROCESSING TO BUILD DTM

Wioleta Błaszczak-Bąk, Artur Janowski

ABSTRACT

Light Detection and Ranging (LiDAR) is a sensing technology which has application in building the Digital Terrain Model (DTM). A point cloud generated from laser scanning makes up the so-called large dataset, which is difficult and sometimes even impossible to use directly. Import of LiDAR point cloud into appropriate software and its processing is time-consuming and demands high computing power. Therefore, it is advisable to optimize the volume of observation results which make up the point cloud.

The following paper presents operation of a modified algorithm for optimization of points' number in a large dataset *[Błaszczak W., 2006]*. The optimization involves reduction and uses existing cartographic generalization methods.

The optimized dataset was filtered, and during the process the points representing the terrain were separated from data representing non-ground elements. Filtration was carried out with the application of a proposed new method including trend line in search belts, and the laser power used to register points.

The optimized and filtered data set was then used to build a DTM.

The results obtained encourage further detailed study of theoretical and empirical character.

1. INTRODUCTION

LiDAR is a sensing technology which has application wherever accurate and quickly accessible data regarding terrain configuration, structure of buildings, and possible danger of natural disasters is needed. Point cloud obtained from an airborne laser scan is used primarily to build a DTM.

The process of LiDAR data processing to build a DTM can be presented in the form of the following flowchart (Fig. 1):

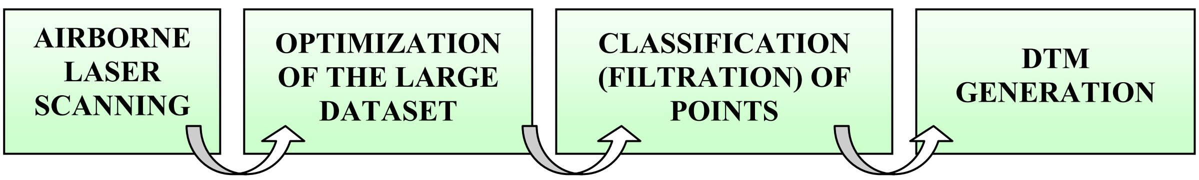


Fig. 1. General Outline of LiDAR Data Processing (Source: Author's Analysis).

During the airborne laser scanning a large dataset is collected $\Omega_{LiDAR}\{P_i(x_i, y_i, z_i), np_i\}$, where $i = 1, 2, ..., k$, k – the number of points in the original dataset from the survey. For each observed point, additional information is also obtained, one regarding the intensity of pulse backscattering from the surveyed object (np_i). The registration of energy value

reflected from the object is of great importance in case of automatic classification of points.
The *point cloud* obtained from laser scanning constitutes the so-called *large dataset,* which is difficult, and sometimes even impossible to use directly, therefore the next stage of LiDAR data processing to build a DTM is the optimization of the large dataset. It involves decreasing the number of observed points in the dataset, which is achieved through reduction based on known cartographic generalization methods. This results in recording of real coordinates of points in the reduced dataset.
During the LiDAR point cloud processing, apart from the optimization issue, there is one of data classification. Classification involves the choice of those points from the reduced dataset, which represent the results of terrain survey. The selected points are written in the dataset representing the terrain, while the points representing the non-ground elements are stored in the set of points representing the non-ground elements.
The last stage of LiDAR data processing is the generation of DTM on the basis of data collected in the dataset representing the terrain.

2. OPTIMIZING THE DATASET NUMBERS

Optimization involving decreasing the number of data in the large set of survey results can be carried out through reduction or generation *[Borkowski A., Jóźków G., 2006, Błaszczak W., Kamiński W., 2007].*
Reduction is based on original data irregularly distributed in the XY plane. Generation is based on an irregular grid of points (most often a grid of squares) and is preceded by proper processing (interpolation) of raw data.
While dealing with the problem of redundancy of observations from continuous, automated surveys, it is important that the optimization of large dataset volume meets the following criteria *[Borkowski A., Jóźków G., 2006]*:

- **if possible, it should be conducted on original data,**
- **the algorithm should have good adjustment properties to the local structures of terrain surface,**
- **the algorithm should enable adding extra information a'priori,**
- **due to the volume of laser scanning datasets, the calculation complexity of the algorithm is not without meaning.**

To reduce the point cloud, a modified algorithm for optimization of large surveying results dataset was used in this paper, as proposed by the coauthor in the doctoral thesis on "Optimization of large sets of surveying results feeding the spatial information systems' databases" *[Błaszczak W., 2006]*. Optimization with the use of the algorithm proposed by the author aims at shortening the time of comprehensive processing and analysis of data obtained, to enable the user an efficient generation of a DTM.

2.1. The algorithm for optimization of large datasets components number and its modification for LiDAR Data reduction

Algorithm for reduction of the number of sets proposed in the paper by *[Błaszczak W., 2006]* consists of the following steps:
***Stage I* - Reading of the dataset. During that stage the points with extreme coordinates have to be determined. There are necessary to determine the processing area range.**

***Stage II* - Determining the width of the search belt within which the reduction takes place.**
***Stage III* - Establishment of horizontal search belts in XOY plane (belts parallel to OY axis).**
***Stage IV* - Selection of the method for elimination of points in search belts. The elimination process is carried out in YOZ plane.**
***Stage V* - Application of the method selected for reduction of points in all search belts in YOZ plane. The elimination process stops when all search belts have been tasted. Points eliminated from individual search belts are placed in the ZDW_{YOZ} (set of data eliminated in plane YOZ).**
***Stage VI* - Establishment of vertical search belts in XOY plane (belts parallel to OX axis).**
***Stage VII* - Application of the method of points reduction selected at stage IV for reduction of point in all vertical search belts. Verification of points is done in XOZ plane. The points eliminated are recorded in the ZDW_{XOZ} (set of data eliminated in plane XOZ). After testing all vertical search belts the elimination process is completed. The eliminated points recorded in sets ZDW_{YOZ} and ZDW_{XOZ} are finally recorded in set ZDW.**
***Stage VIII* represents completing the initial processing and writing the resulting set.**

By dint of fact the LiDAR measurement is conducted along survey line by given length, resulted from laser scanning angle and height of flight, during the algorithm executing there is not necessity of creating the search belts. Generalization method can be applied in original belts, which were created during LiDAR measurement. These areas are called now measuring belts. Survey point affiliation to certain area is defined in text file. This file is the result of the LiDAR measurement. Information is presented in table *nrP(N)*, where *NrP - no. of search belt, N - number of belts*. The possibility of point elimination in original LiDAR measurement areas significantly simplified the algorithm execution, what leads to following, stages:
***Stage I* - Reading of the dataset LiDAR (set $\Omega_{LiDAR}\{P_i(x_i, y_i, z_i), np_i\}$).**
***Stage II* - Selection of the method for generalization.**
***Stage III* - Application of the method selected for reduction of points in all original measuring belts.**
***Stage IV* – Completion of optimization processing steps and writing the resulting dataset $\Omega_R\{P_i(x_i, y_i, z_i), np_i\}$ (where Ω_R is the dataset after optimization, $i = 1r, 2r, ..., m$, m – the number of point in reduced dataset) to the local software.**

3. LiDAR DATA FILTRATION

Filtration (classification) of laser data is a procedure of elimination of points referring to non-ground elements from the observed point cloud. It allows determining the topographic surface, i.e. generating the DTM *[Marmol U., Jachimski J., 2004]*.
The external sources on the topic, mention several LiDAR data filtration methods, e.g.:

a) **algorithm based on digital image processing: morphological filters [*Zhang K. and others, 2002, Vosselman G., 2001*], gradient methods [*Hyyppä and others, 2002, Wack, Wimmer, 2002*], active surface modeling [*Elmqvist, 2002*],**

b) algorithm based on interpolation methods: linear prediction [*Pfeifer and others, 2001*], spline interpolation [*Brovelli and others, 2002*], active TIN method [*Axelsson, 2000*].

The existing methods have their limitations, related to the incorrect filtration of points in areas of complex lay of the land and terrain use, therefore the study to construct new filtration algorithms is still under way. All algorithms require a significant, interactive participation of an operator.

3.1. Proposed data filtration algorithm

The classification algorithm proposed in this paper involves testing of observed points separately, in each search belt. On the basis of the assumed criteria, the user decides to which dataset to qualify a given surveying point.
Generally, the process of assignment of points to datasets can be presented as follows:

$$\Omega_R\{P_i(x_i, y_i, z_i), np_i\}$$

$$\Omega_{PNMT}\{P_i(x_i, y_i, z_i), np_i\} \qquad \Omega_{PSZ}\{P_i(x_i, y_i, z_i), np_i\}$$

where:
Ω_{PNMT} – set of points representing the terrain,
i – succeeding point in the dataset representing the terrain
($i = 1, 2, ..., w$; where w – number of points in the dataset Ω_{PNMT})

where:
Ω_{PSZ} – set of points representing non-ground elements,
i – succeeding point in the dataset representing the non-ground elements
($i = 1, 2, ..., z$; where z – number of points in the dataset Ω_{PSZ})

It is important to stress that the Ω_{PNMT} dataset will include, apart from the points representing the terrain, the non-ground elements, which do not interfere with the terrain configuration. These may include e.g. manhole openings, gas and water valves and similar non-ground elements, which are permanent elements of the ground level and their height corresponds to the elevation in a given spot.

The analysis of so-called auxiliary (ancillary) materials is necessary to realize the algorithm correctly. The most important ones are:

a) materials regarding the data source,
b) information regarding the used laser scanner, e.g. scanning belt width, angle, length of scanning belts, distance between points in the scanning belt (scanning resolution),
c) information regarding the land development and configuration of the measured terrain (e.g. mean absolute elevation of terrain),
d) air photographs, supporting the process of data classification, they allow the assessment of the generated DTM.

The process of data classification starts from importing the optimized dataset $\Omega_R\{P_i(x_i, y_i, z_i), np_i\}$. The set is divided into search belts. The algorithm analyses each

belt. An example of a search belt resulting from scanning with constant scanning angle is presented in Fig. 2.

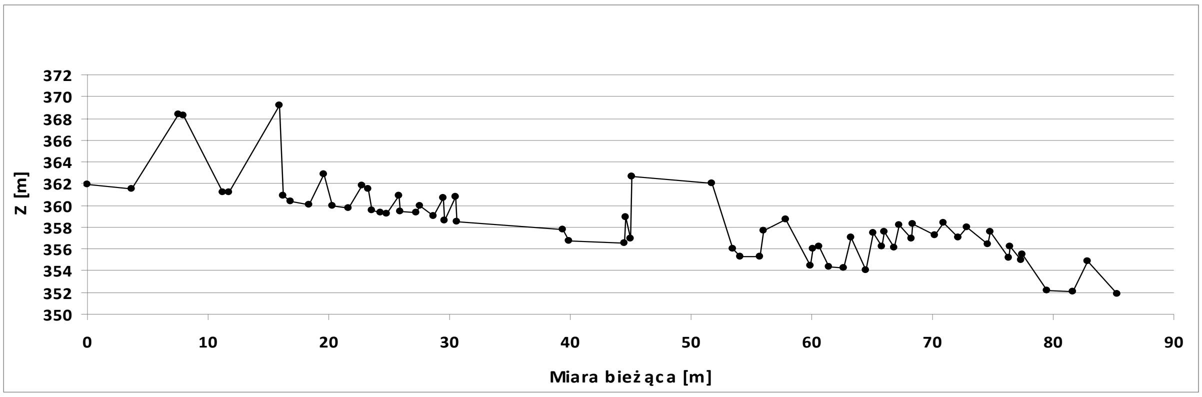


Fig. 2. Example of Search Belt from Laser Scanning (Source: Author's Own Work).

At the first stage of the algorithm, based on points in the search belt, the trend line is determined. It allows the first stage of points' classification and assignment of points to Ω_{PSZ} set. The determined trend line and the first stage of classification are presented in Fig. 3.

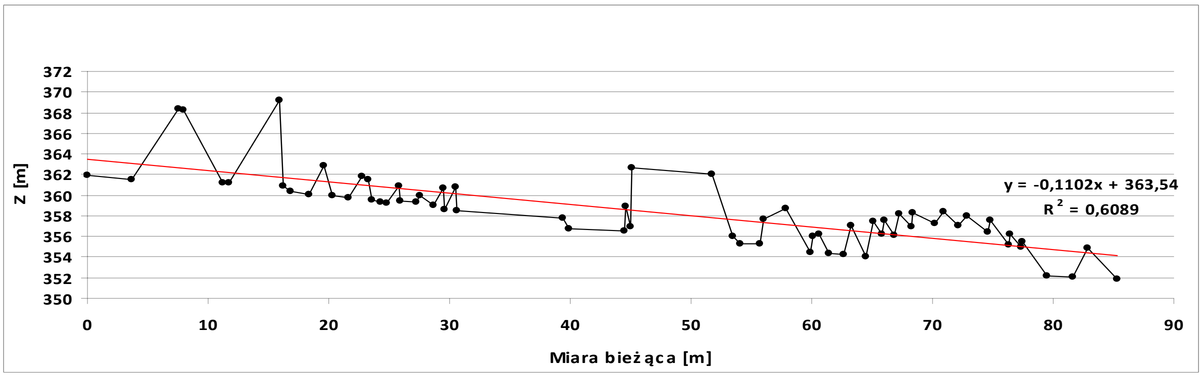


Fig. 3. Introduction of the First Trend-line (Source: Author's Own Work).

The points located above the trend line are assigned to the set Ω_{PSZ}, while the points below the trend line undergo the second stage of classification. For this purpose, the second trend line is generated based on the points located below the trend line. This stage is presented in Fig. 4.

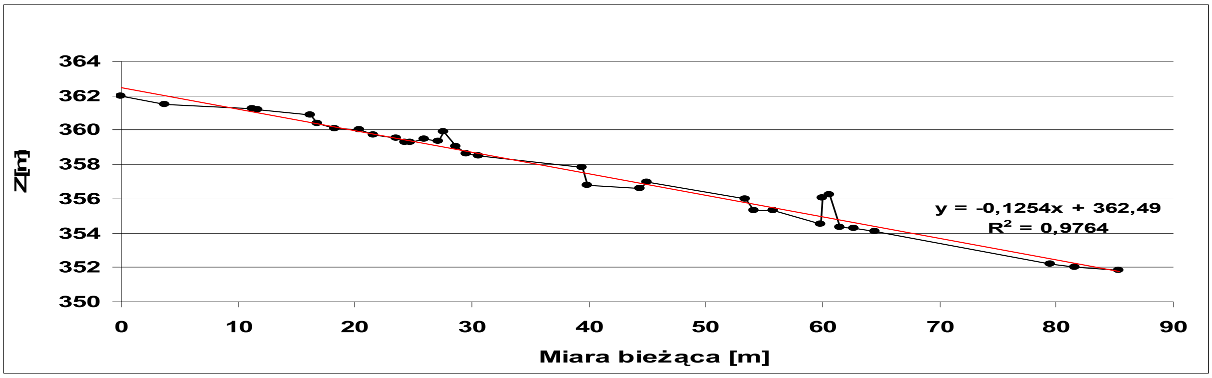


Fig. 4. Introduction of the Second Trend line (Source: Author's Own Work).

The process of classification of points for the second trend line begins with the determination of an average height between points $z_{śr}$, which take part in the second stage of classification. In the presented example, $z_{śr}$ = 0,306 m. The next stage is the calculation of d_i distance points, located above the second trend line and comparison with the average relative height between points. If di > $z_{śr}$, the point is assigned to set Ω_{PSZ}, if $d_i < z_{śr}$, the point is assigned to set Ω_{PNMT}. Additionally, the laser power *(np_i)* is checked for the starting point and compared with the previously determined ranges of laser power for the terrain (range: $np_{NMTi.}$ $<a_1,b_1>$) and non-ground elements (range: $np_{SZi.}$ $<a_2,b_2>$). The points which are located below the trend line are automatically recorded in set Ω_{PNMT}.

At the subsequent stages of the algorithm, the succeeding points of the search belt are being tested. The process is finished when all points in the search belt have been tested. Once all points in the first search belt have been tested, succeeding search belts are tested. When the algorithm operation ends, we obtain two sets of points:
$\Omega_{PNMT}\{P_i(x_i, y_i, z_i), np_i\}$ i $\Omega_{PSZ}\{P_i(x_i, y_i, z_i), np_i\}$.

4. EXAMPLE OF PRACTICAL APPLICATION

The algorithm for optimization of large surveying results dataset was applied to reduce the number of observations in a point cloud from surveying of a section of a newly designed road in Bielsko-Biała. The original dataset is presented in Fig. 5.

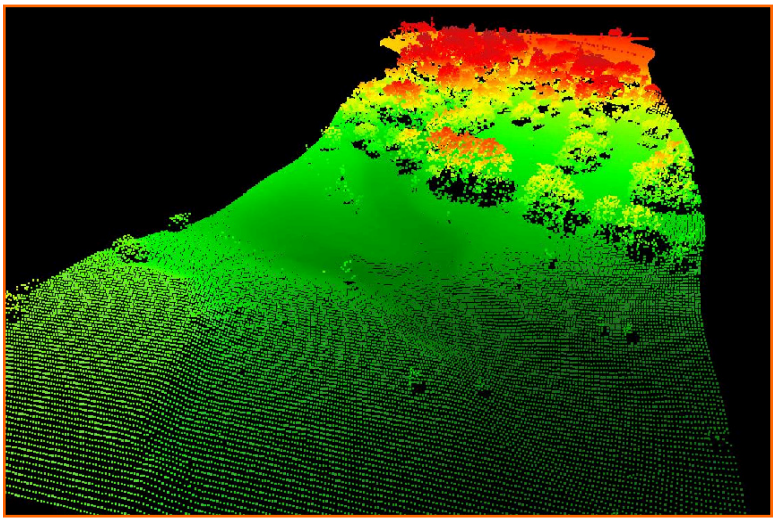


Fig.5. LiDAR Point Cloud (Source: Visimind Ltd).

The tested large dataset comes from Visimind surveying system, which comprises of Riegl LMS-Q240 laser, GPS Topcon, IMU, and digital cameras. The laser scanning angle equaled 60 degrees, with resolution of 10 000 Hz. The scan was made during a helicopter flight with the speed of around 50 km/h at the height of around 70 m. The subset of LiDAR data used for digital tests contains around 1 400 000 points.

4.1. LiDAR dataset optimization

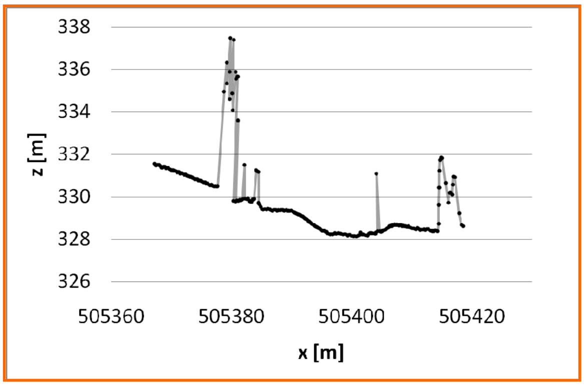
Stage I of the algorithm realization involved input of point cloud of the dataset $\Omega_{LiDAR}\{P_i(x_i, y_i, z_i), np_i$. The cloud was written down in the form of a text file, with tables

including the following: *no. of search belt nrP(N), no. of observation point in the belt nr(M), coordinate X(M) in 1992 coordinates system, coordinate Y(M) in 1992 coordinates system, normal height H, laser power np(M)*, where *N* denotes the number of belts, *M* stands for the number of the points in the belt.

In the examined LiDAR point cloud, the coordinate $N = 0, 1, 2, ..., 5380$, while $M = 0, 1, 2, ..., 248$. The length of search belt equals around 81 m. The average distance between points in the belt is around 0.40 m.

At the Stage II of the algorithm, the cartographic generalization method was determined. The Douglas-Peucker method was chosen to reduce the number of points in the dataset.

Stage III of the algorithm – the chosen generalization method was applied in the respective search belts. To realize the algorithm, the tolerance range of 0.40 m was assumed. This value was determined on the basis of the average distance between points in the search belt. The distribution of points in an example search belt before and after reduction with the Douglas-Peucker method is presented in Fig. 6.



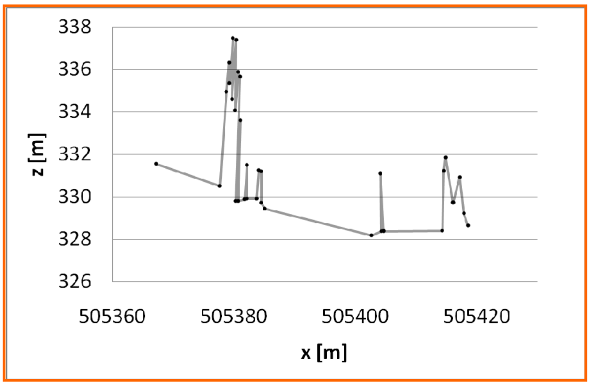


Fig. 6a) Distribution of Points in an Example Original Search Belt; b) Distribution of Points in an Example Search Belt after Reduction (Source: Author's Analysis).

At Stage IV the output file was recorded by the local software. The optimized dataset has capacity almost 11 times lower than the original dataset and contains about 120 000 points. The points eliminated during the realization of the algorithm were assigned to dataset ZDW.

The application of the optimization algorithm in Ω_{LiDAR} set allowed further, more effective and less time-consuming processing of the point cloud. At the following stage of LiDAR data processing, namely the classification, an optimized set Ω_R will be used.

4.2. LiDAR data filtration

Data filtration was conducted on the optimized dataset Ω_R .

Additional materials at disposal, during determination of the dataset representing the terrain and set of points representing the non-ground elements, were as follows:

a) **data source: Visimind,**
b) **laser scan angle: 60^{o},**
c) **scanning frequency: 1000 Hz,**
d) **airplane flight height: around 70 m,**
e) **type of terrain: gently sloping,**

f) land development: agricultural with forestation,
g) air photographs.

In order to realize the algorithm correctly, two ranges of laser power were assumed. For the terrain the range <0,0706; 0,1500> was determined, while for the non-ground elements the range was (0,1500; 0,3059>.
Fig. 7 below presents an example search belt after filtration.

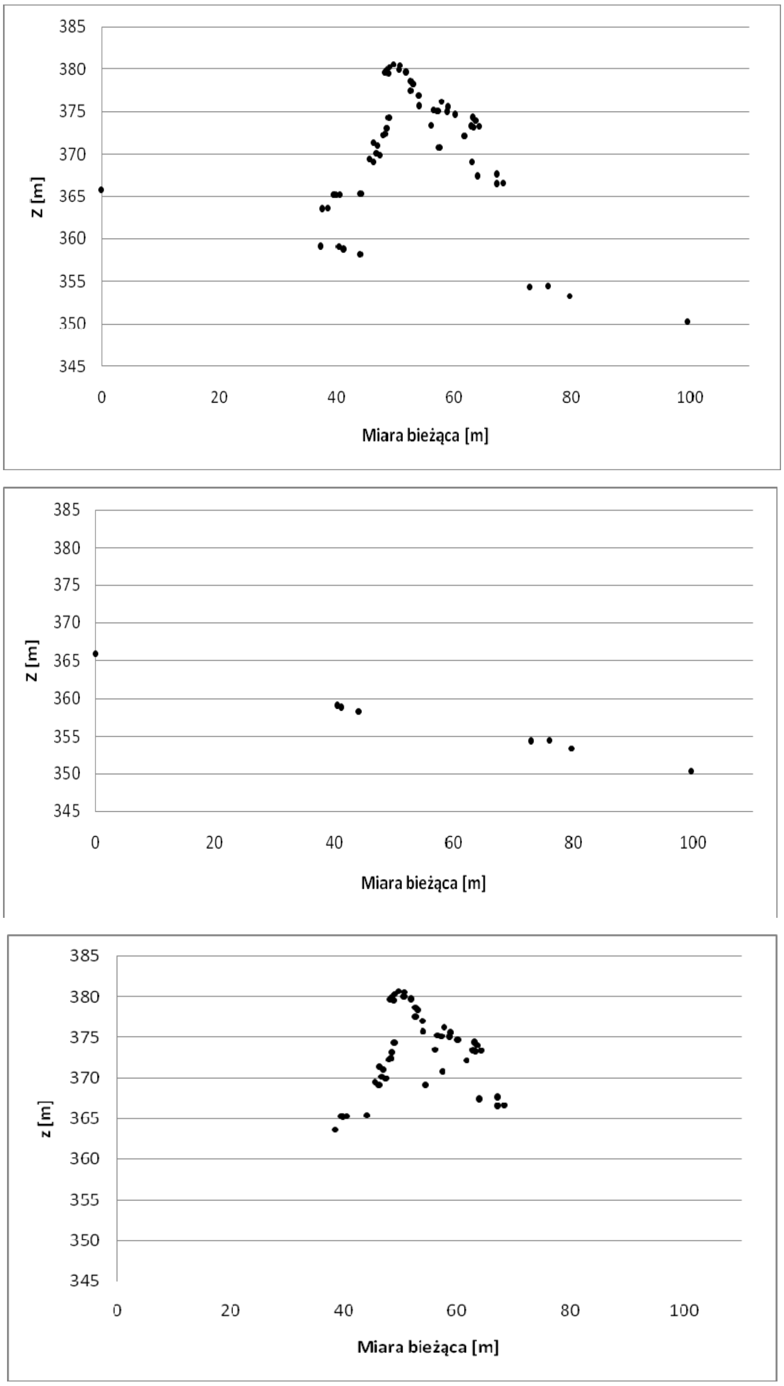


Fig. 7 a) Distribution of Points in an Example Search Belt after Reduction; b) Distribution of Points in an Example Search Belt in the Dataset Ω_{PNMT}; c) Distribution of Points in an Example Search Belt in the Dataset Ω_{PSZ} (Source: Author's Own Work).

5. NMT GENERATION

On the basis of the Ω_{PNMT} dataset, the NMT presented in Fig. 8 was generated. The size of grid cell of 2m was assumed, and the interpolation was conducted with Kriging method.

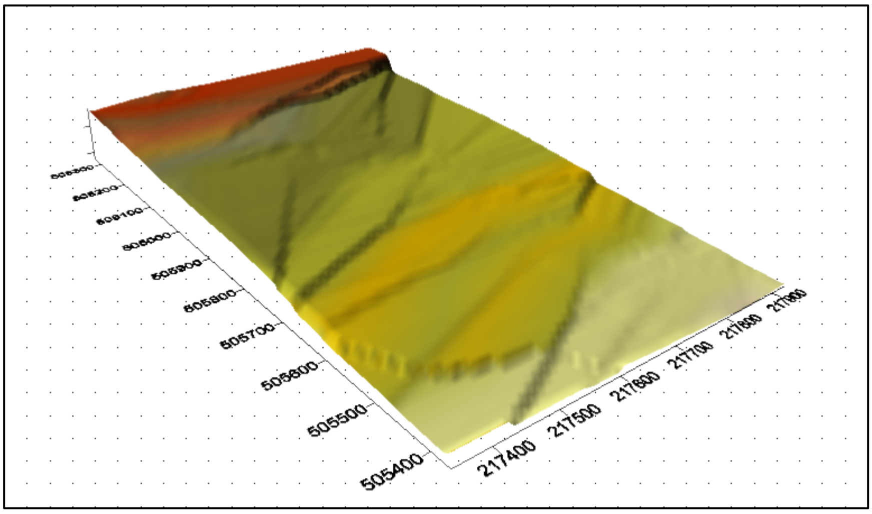


Fig. 8. DTM Generated in Surfer v.8 (Source: Author's Own Work).

The generated DTM can be used in architectural project works on the examined area.

6. CONCLUSION

The result of study conducted in this paper, allows to state that LiDAR data filtration is definitely a prospective subject. The process of optimizing the dataset obtained through the airborne laser scanning, which can precede the filtration process, greatly facilitates data classification. This leads to a more effective and quicker preparation of data to build a DTM without loss of information necessary to realize the task correctly.

It is important to mention that the success of the proposed algorithm of LiDAR data processing will depend to a great degree on terrain type. As there are no fully automated data filtration algorithms, one should expect possible filtration errors. In case of only slight changes in slope angle, one can assume that the errors will be considerably smaller than for complex terrain configuration.

REFERENCES

Axelsson P., 2000: DEM generation from laser scanner data using adaptive TIN models. International Archives of Photogrammetry and Remote Sensing Vol. XXXIII/4B, Amsterdam.

Błaszczak W., 2006: Optymalizacja dużych zbiorów wyników pomiaru zasilający bazy danych systemów informacji przestrzennej. Praca doktorska. Olsztyn.

Błaszczak W., Kamiński W., 2007: Data number reduction in measurement results set using optimization algorithm. FIG Working Week 2007. Strategic Integration of Surveying Services. The Hong Kong Institute of Surveyors. Hong Kong. CD-ROM.

Błaszczak W., Kamiński W., 2007: Optymalizacja dużych zbiorów wyników pomiaru zasilających bazy danych systemów informacji przestrzennej. Przegląd Geodezyjny nr 6/2007, str. 7-9.

Brovelli M. A., Cannata M., Longoni U. M., 2002: Managing and processing LIDAR data within GRASS. Proccedings of the Open source GIS – GRASS user conference, Trento.

Hyyppä J., Pyssalo U., Hyyppä H., Samberg A., 2002: Elevation accuracy of laser scanning – derived digital terrain and target models in forest environment. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIV/ 3A, Graz.

Marmol U., Jachimski J., 2004: A FFT based method of filtering airborne laser scanner data. International Archives of Photogrammetry and Remote Sensing Vol. XXXIII/3, s. 1147-1152, Istanbul.

Vosselman G., 2001: Adjustment and filtering of raw laser altimetry data. OEEPE Workshop on Airborne Laserscanning and Interferometric SAR for Detailed Digital Elevation Models, Stockholm.

Wack R., Wimmer A., 2002: Digital terrain models from airborne laser scanner data – a grid based approach. International Archives of Photogrammetry and Remote Sensing, Vol. XXXIV / 3B, Graz.