

# FILTERING PROCESS OF LIDAR DATA

BADEA Dragos<sup>a</sup>, JACOBSEN Karsten<sup>b</sup>

<sup>a</sup>UTCB, Technical University of Civil Engineering Bucharest, 124 Lacul Tei Bd., Bucharest, Romania -  
badea\_dragos@yahoo.com

<sup>b</sup>IPI, Institute for Photogrammetry and GeoInformation, Institute Leibniz University of Hannover, 1 Nienburger Str.  
D30167 Hannover, Germany, jacobsen@ipi.uni-hannover.de

WG I/5

**KEY WORDS:** filtering, lidar, break-lines, clasification, tin, DHM, DSM,

## ABSTRACT:

This paper describes the processing method of the LiDAR data by using the feature characteristic derived from height and position information. For the analysis of Digital Height Model (DHM) resulted from LIDAR data collection, many researches have been made, the main objectif being the similitude between digital representation and real terrain situation. Many problems occur when trying to eliminate points not belonging to bare terrain surface, but on man made objects and vegetation. Points located outside the bare terrain surface can be eliminated through a lot of filtering methods, but a special attention should be paid in order not to eliminate real terrain located points, which can be easily mistaken as located on man made objects. As an example, sudden change of terrain slope can be mistaken for a building roof. Those objects not belonging to the bare earth can be eliminated through different filter methods. In a flat area, an object can be detected by an algorithm which analyses the height of the points in relation to the surrounding area. The same algorithm is not anymore suited to filter off-terrain objects from the area with the same threshold. You can have the surprise that this filter will eliminate points from bare earth surface. This paper-work bring you a algorithm for filtering the Digital Surface Model to achieve a more accurate Digital Height Model. The purpose of this research is to extract DHM (Digital Height Model) from LiDAR data. DSM is the model that includes existing terrain and non-surface objects (e.g. trees, buildings, cars etc.). Various applications need the DHM - Digital Height Model. This research proposes another method that processes the raw data by using the concept of combining the classification and filtering processes into one, also using automatic detected break lines with ArcInfo. The methods of filtering and generation of DHM I use, were developed, implemented in software and improved in time at Institute of Photogrammetry and GeoInformation at the University of Hannover (IPI).

## 1. Data colector - LIDAR

A pulsed laser ranging system is mounted in an aircraft equipped with a precise kinematic GPS receiver and an Inertial Navigation System (INS). The signal is sent towards the earth where it is reflected by a feature back towards the aircraft. A receiver then captures the return pulse. Using accurate timing, the distance to the feature can be measured. By knowing a speed of the light and the time the signal takes to travel from the aircraft to the object and back to the aircraft, the distances can be computed. Using a rotating mirror inside the laser transmitter, the laser pulses can be made to sweep through an angle, tracing out a line on the ground. By reversing the direction of rotation at a selected angular interval, the laser pulses can be made to scan back and forth along a line.

Errors in the location and orientation of the aircraft, the beam director angle, atmospheric refraction model and several other sources degrade the co-ordinates of the surface point to 5 to 10 centimetres. An accuracy validation study showed that Lidar has the vertical accuracy of 10-20 centimetres and the horizontal accuracy of approximately 1 meter.

Among Lidar technology advantages, these systems afford the opportunity to collect terrain data about steep slopes and shadowed and inaccessible areas (such as large mud flats and tidal areas). Following the initial post processing and positional assurance, the Lidar data are filtered for noise removal and prepared as a file of x,y,z points.

After that the point coordinate X, Y, Z is obtained by analyzing the position, attitude, distance and projection transformation. Since this data includes many Non Surface Objects such as buildings, trees and cars, the classification is needed in various application fields. This classification is used in the filtering process of LiDAR data.

## 2. From Digital Surface Model to Digital Height Model

A method for the generation of DHM from airborne Lidar data is presented which was developed and improved at IPI. The method distinguishes itself by using the filter in the same time with interpolation.

The original data obtained by Airborne Laser Scanner express the surface of ground objects, not only the ground surface but also trees and roofs of buildings. These data are called digital surface model (DSM). It is necessary to distinguish these ground objects and to create a digital height model (DHM) that expresses the ground elevation by removing trees and buildings from DSM. This process is called "filtering." The manual refinement is time consuming. The strategy for automatic filtering is based on a combination of geometric conditions together with Linear Prediction. The filtering was applied to photogrammetric acquired data via automatic image matching.

RASCOR is a program developed for automatic improvement of digital surface models to digital elevation models (raster data set).

Program RASCOR can analyze, improve, smooth and interpolate a digital elevation model (DEM) which may be created by automatic image matching or laser scanning (LIDAR) in an equal spacing arrangement. The identification of points not located on the solid ground but on topographic features like vegetation and buildings is possible by a minimal and maximal height in the area, by maximal height differences between neighbored points, by a sudden change of the height level, by a linear or polynomial interpolation in X- and Y-direction, by a minimal and maximal height difference against a local tilted plane or polynomial surface and a local prediction (least squares interpolation) based on the tilted plane or polynomial surface.

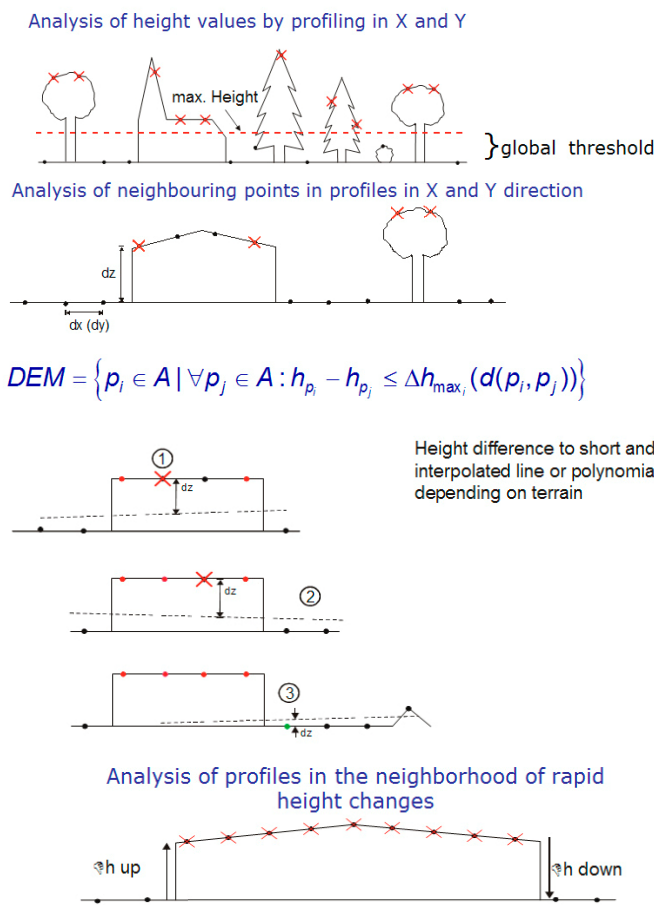


Figure 1. Different terrain analysis.

The final results can be filtered (smoothened) in relation to an inclined moving plane or polynomial surface fitted to the neighbored.

The required parameters can be automatically determined by an analysis of the DEM based on a simple characterization of the area as homogenous or not and smooth or undulated or mountainous during main program dialog.

RASCOR can respect break lines during data handling. The break lines can be also introduced as an ASCII file containing the coordinates along the lines. They also can be manually

imputed or added in the iterative filtering process. In this case break lines file was automatically generated with ArcInfo.

The interpolated surface in the point  $P_i$  is given by:

$$u_i = \underline{c}^T \underline{C}^{-1} \underline{z} \quad (1)$$

$$\underline{c}^T = [C(P_i P_1) \quad C(P_i P_2) \quad \dots \quad C(P_i P_n)]$$

$$\underline{C} = \begin{bmatrix} V_{zz} & C(P_1 P_2) & \dots & C(P_1 P_n) \\ C(P_2 P_1) & V_{zz} & \dots & C(P_2 P_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(P_n P_1) & C(P_n P_2) & \dots & V_{zz} \end{bmatrix}$$

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \quad C(P_i P_k) = C(0) e^{-A \left(\frac{PR}{B}\right)^2}$$

- $u_i$  predicted value
- $\underline{c}$  covariances between point to be interpolated and measurements
- $\underline{C}$  covariance matrix
- $\underline{z}$  vector of centered measurements

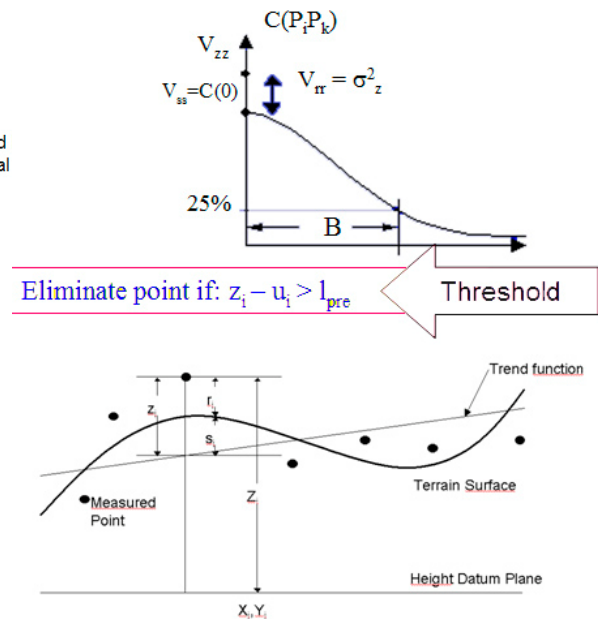


Figure 2. Interpolation method

Best results have been obtained with the 'LoG-Operator' within HALCON, which uses the Laplacian-Operator  $\Delta g(x,y)$  and a selectable smoothing  $\sigma$  of the Gauss-function.

Implementation:

$$\Delta g(x,y) = \frac{\partial^2 g(x,y)}{\partial x^2} + \frac{\partial^2 g(x,y)}{\partial y^2}$$

The derivatives of the LoG are approximated by derivatives of the Gauss-function  $G\sigma(x,y)$

$$G_{\sigma}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$$

and

$$\Delta G_{\sigma}(x,y) = \frac{1}{2\pi\sigma^4} \left(\frac{x^2+y^2}{2\sigma^2} - 1\right) \left[\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\right]$$

This results in a detection of an ideal edge having a maxima and a minima and a zero crossing (steepest slope of the edge).

Top (upper) edge of dike could not be detected because segmentation via thresholds failed. Therefore the mean curvature  $H$  is determined from the derivatives of the Gauss-function:

$$G_{\sigma}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$$

$$H = \frac{a-b+c}{d}$$

$$a = \left(1 + \frac{\partial g(x,y)^2}{\partial x}\right) \cdot \frac{\partial^2 g(x,y)}{\partial y^2}$$

$$b = 2 \frac{\partial g(x,y)}{\partial x} \cdot \frac{\partial g(x,y)}{\partial y} \cdot \frac{\partial^2 g(x,y)}{\partial y \partial x}$$

$$c = \left(1 + \frac{\partial g(x,y)^2}{\partial y}\right) \cdot \frac{\partial^2 g(x,y)}{\partial x^2}$$

$$d = \left(1 + \frac{\partial g(x,y)^2}{\partial x} + \frac{\partial g(x,y)^2}{\partial y}\right)^{\frac{3}{2}}$$

A break line will avoid an elimination at locations with rapid change of the inclination like a dam.

Program RASCOR is using a sequence of different methods for the filtering of a DSM. Only data sets with raster arrangement are accepted.

RASCOR starts with an analysis of the height distribution itself. This methods requires flat areas, it does not work in rolling and mountainous terrain. It is followed by an analysis of the height differences of neighboured points. The accepted height limit of neighboured points is depending upon the slope and the random errors. With this method only small objects and the boundary of larger elements can be eliminated, but it is still very efficient.

Even large buildings can be found by a sudden change of the elevation in a profile to a higher level and a later corresponding change down if no vegetation is located directly beside the buildings. This method is used for laser scanning, but it is not optimal for DEMs determined by automatic image matching where the buildings are looking more like hills.

Other larger objects not belonging to the bare ground are identified by a moving local profile analysis; at first shorter and after this longer profiles are used. The required length of the moving local profiles is identified by an analysis of a sequence of shorter up to longer profiles. In flat areas the individual

Height values are checked against the mean value of the local moving profile, in rolling areas a linear regression is used, in mountainous areas polynomials have to be used. It will be combined with data snooping taken care about a not even point distribution caused by previously eliminated points. All these methods are applied in X- and Y-direction. Elements which have not been removed by this sequence of tests are analysed by moving surfaces which may be plane, inclined or polynomial. The size of the moving surfaces is identified by the program itself by checking the data set with a sequence of cells with different size. As final test a local prediction can be used, but it is usually only finding few points not belonging to the surface after the described sequence of tests.

In the case of the check for height differences of directly neighboured points, the upper point will be eliminated if the tolerance limit will be exceeded. The other methods are using a weight factor for points located below the reference defined by the neighboured points. This will keep points located in a ditch or cutting in the data set. Usually points determined by laser scanning do not have blunders causing a location below the true position, but this may happen in the case of a DSM determined by automatic image matching, justifying a weight factor.

In forest areas at first only the trees are removed by the program, smaller vegetation is remaining, so a second iteration is necessary. A second iteration in other cases may remove also terrain points leading to a more generalised DEM. This may be useful for the generation of contour lines, but it is not optimal for the correct description of the terrain.

### 2.1. Break lines

Linear features that define critical changes (natural or manmade) in topographical shape. Breaklines define and control surface behavior in terms of smoothness and continuity. They have a significant effect in terms of describing surface behavior when incorporated in a surface model such as a triangulated irregular network (TIN).

The most important reason why break lines should be located in global object reconstruction is the demand to keep the number of unknown geometric, DEM parameters low so that large image areas could be processed simultaneously and as fast as possible. As in the least squares adjustment the size and the structure of the normal equation matrix depend on the unknown quantities, and as the object surface elements can be eliminated in the matching, the amount of unknowns depends mostly on the number of DEM grid points. Beside the largest errors of the matching occur due to break lines when a continuous model is used in object surface reconstruction. Thus, if a continuous model is used without considering break lines, the result of the image matching is not reliable especially at break line locations. These reasons imply that in global object reconstruction break lines should be detected at first so that break line areas could be better modelled.

Breaklines were obtained automaticaly with ArcInfo and introduced in the filtering process.

The breaklines were produced using the Arc/INFO GENERATE command. The INPUT file was a .GEN file produced with an script which contained DTM point coordinates. All of these data were extracted from the original DSM files from Lidar.

Break lines have been obtained with ArcInfo by computing a TIN. A triangulated irregular network data model (TIN) is an efficient way for representing continuous surfaces as a series of linked triangles. A tin is formed by nodes, triangles and edges. Nodes are locations defined by x, y and z values from which a tin is constructed. Triangles are formed by connecting each node with its neighbors according to the Delaunay criterion: all sample points are connected with their two nearest neighbors to form triangles (by using this method the triangles are as equi-angular as possible, any point on the surface is as close as possible to a node, and the triangulation results independent of the order the points are processed). Edges are the sides of triangles.

Points used in defining the tin are called mass points. Areas of constant elevation, such as water surfaces are called exclusion polygons. Finally, lines such as streams and shorelines are called breaklines. Breaklines can be either hard or soft. Hard features are things like roads, streams, and shorelines which indicate a significant break in slope. Soft features are things like ridgelines on rolling hills. Ridges like these do not represent distinct breaks in slope but since they separate watersheds you might like to maintain in the triangulation. Newly automatical extracted break lines will be introduced to filtering algorithm.

For the interpolation the “moving plane” method had been used. The plane is stopping at the break line, and will not eliminate the points beyond it..

The laser scanner data are a point cloud without structural information. A qualified terrain model, on the other hand, exists in the inclusion of structure lines, especially break lines. A terrain model, integrating a close down grid (raster data) and structure lines (vector data) , is called hybrid DHM. The raster data as well as the vector data is smoothed, the hybrid DHM shows discontinuities in the first derivation at the break lines.

The solution found, combines filtering and interpolation of terrain. Is called a robust interpolation or robust linear prediction. The algorithm is embedded in a hierarchical approach, however, the filtering of the laser scanner data on one level will be described first. In this Algorithm a rough approximation of the surface is computed first. Next, the residuals, i.e. the oriented distances from the surface to the measured points are computed. Each (z) measurement is given a weight according to its distance value, which is the parameter of a weight function. The surface is then recomputed under the consideration of the weights. A point with a height weight will attract the surface, resulting in a small residual at this point, whereas a point that has been assigned a low weight will have little influence to the figure of the surface. If an oriented distance is above certain value, the point is classified as off-terrain point and eliminated completely from this surface interpolation. This process of weight iteration is repeated until all gross are eliminated or a maximum number of iterations is reached. Program RASCOR use maximum 2 iterations. The reason for this limitation is that more iterations will result in a very smooth earth surface due to the risk of elimination of too many points.

For the interpolation we use linear prediction. In this method the classification and DHM generation are performed in one step, there is no assumption that the terrain is horizontal. It is applied patch wise to the data, which result is an adaptive setting of the shift origin of the weight function. Furthermore, the base are determined for each patch separately, too. The

process yields a smooth surface, that means the accidental (random) measurement errors have also been filtered.

However, the algorithm relies on a “good mixture” of ground and off-terrain (vegetation) points, which can also be seen as a high frequency of change from change from ground to vegetation points. This is necessary for a reliable determination of the shift value for the origin of the weight function. In this high frequency is not given, we need to provide the input data in a suitable form. This can be achieved by inserting the robust linear prediction in a hierarchic environment.



Figure 3. Mosaic for the study area

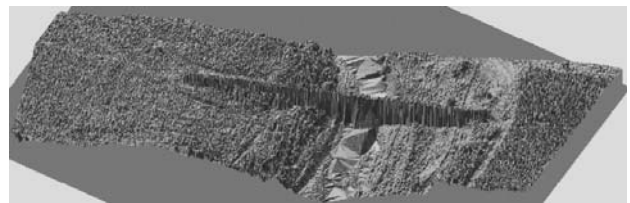


Figure 4 DSM – not filtered



Figure 5 DHM – filtered (with break lines)

### 3. Conclusions

- Automatic filtering of a DSM → DEM without the consideration of the different class dependant height structures (global view) shows unsatisfactory results
- Especially areas of settlement, large buildings, forests and dam-like structures require local parameter settings
- Grouping of classes like:
  - Settlement and large buildings
  - Water courses (rivers, canals,...)
  - Highways and dams
  - Fields, crops, areas of low vegetation is possible with no considerable accuracy loss.
- Breaklines should be included especially for dams, highways watercourses and bridges

Very good results were achieved when using break lines for filtering the DSM to obtain DHM. Less points were removed from initial data in comparison with case in which break lines are not used.

The filtering and interpolation works very fast and in a simple to use interface. The results obtained with the programs are satisfactory when we are taking care of the above concluded listed points.

The operator must have a image of the area or other type of data, in order to set correctly the parameters for program RASCOR. Else, the results may be wrong just because of not knowing some details which are very important for the filter process.

As a general conclusions, I must say that a Digital Terrain Model may be improved depending on each one imagination for mathematical methods and other data or parameters joined with pure Lidar data, the procedure of filtering will remain a big topic of interest because it can never be left to complete automatisation. There will always be required the human operator to judge the result and by nature, humans will always try to improve it.

#### **References and/or Selected Bibliography**

Jacobsen, K. 2001, PC-Based Digital Photogrammetry, UN/COSPAR ESA –Workshop on Data Analysis and Image Processing Techniques, Damascus, 2001 ; volume 13 of “Seminars of the UN Programme of Space Applications, selected Papers from Activities held in 2001”

Jacobsen, K. 2003, RASCOR – manual, Institute for Photogrammetry and Engineering Survey ,University of Hannover, Germany

Jacobsen, K., Lohmann, P., 2003, SEGMENTED FILTERING OF LASER SCANNER DSMS , ISPRS WG III/3 workshop „3-D reconstruction from airborne laserscanning and InSAR data“, Dresden 2003

Passini R., Betzner D., Jacobsen K., 2002, Filtering of Digital Elevations Models, ASPRS annual convention, Washington 2002

K. Kraus and N. Pfeifer 2002, ADVANCED DTM GENERATION FROM LIDAR DATA  
Institute of Photogrammetry and Remote Sensing  
Vienna University of Technology  
A-1040 Vienna, Austria

Ilkka Korpela 3D data capture for DEM/DTM/DSM production - an introduction to photogrammetric methods and ranging laser  
For course Y196, November 2000 University Of Helsinki

E.P. Baltsavias , 1999 , Airborne laser scanning : existing systems and firms and other resources . Institute of Geodesy and Photogrammetry, ETH-Hoenggeberg, Zurich Switzerland .  
ISPRS Journal of Photogrammetry and Remote sensing 54 pag. 164-198

Yukihide Akiyama , 2000, The advantages of high density airborne laser measurement , Spatial Information Department , Working group V/1 , International Rchives of Photogrammetry and Remote Sensing Vol. XXXIII, Part B4, Amsterdam 2000, Aero Asahi Corporation Japan

Krysia Sapeta, LIDAR, Analytical Surveys Inc. (ASI), Colorado Springs, Colorado

Esteban Azagra, 1999, How to create a TIN from a GENERATE input file

Office of Geographic and Environmental Information (MassGIS), 2003 , Digital Orthophoto Topographic Breaklines

#### **Aknoledgements:**

Special thanks to Institute for Photogrammetry and GeoInformation Hannover, Faculty of Geodesy Bucharest and S.C. InterGIS S.R.L. for provided support.

