

Filtering Process of LiDAR Data

*Dragos BADEA, Asist. univ. drd. ing, Technical University of Civil Engineering Bucharest, badea_dragos@yahoo.com,
Karsten JACOBSEN, Dr. ing. Leibniz University of Hannover - Institut for Photogrammetry and GeoInformation, jacobson@ipi.uni-hannover.de*

Abstract: *This paper describes the processing method of the LiDAR data by using the feature characteristic derived from height and position information.*

The most important factor that everyone cares, is the similarity of digital representations with the real terrain surface. The LiDAR data is classified by the neighborhood height effect.

The problem occurs when you try to eliminate off-terrain points like manmade objects and trees. 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. And you have to find a different approach to the problem.

This paper-work demonstrate the benefits brought by introduction of break lines in filtering the Digital Surface Model to achieve a more accurate Digital Height Model.

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).

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.

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 called the filtering process of LiDAR data.

The purpose of this research is to extract DSM (Digital Surface Model) from LiDAR data

Sometimes the raw data itself is called DSM and sometimes not. In this paper, DSM is defined as following to eliminate complications.

DSM = DTM + Non Surface Objects

DTM (Digital Terrain Model)

In other words, DSM is the model that includes existing terrain and non-surface objects (e.g. trees, buildings, cars etc.).

This research proposes a method that processes the raw data itself by using the concept of combining the classification and filtering processes into one, also having the possibility to use semi-automatic detected break lines.

1. Obtaining DHM from Lidar data

Among their 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.

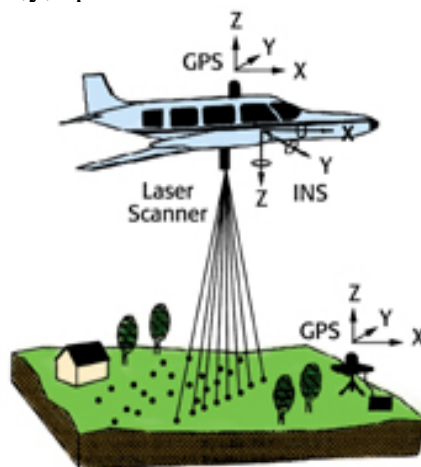


Fig. 1 Theoretical pulse emitted from a Lidar system

The imagery can be used to add break lines to the Lidar data to reveal the terrain more accurately.

2. DHM generation from Lidar data and filtering

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 ALS 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.

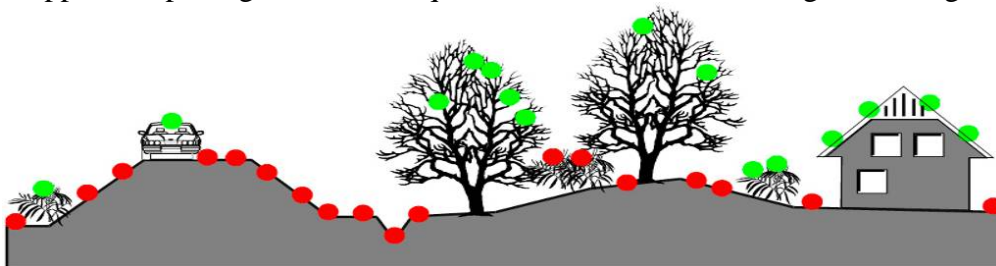


Fig. 2 DSM Yellow points belonging to surface of objects.
DHM Only red points – belonging to bare earth surface.

2.1. Break lines

Digitized lines that define critical changes (natural or manmade) in topographical shape. 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.

Thus, break lines can be detected by computing one ortho-image per aerial input image and the difference (or deviation image in case of more than two input images) ortho-image at the same image pyramid level, and interpreting large differences between the ortho-images as defects in the geometric model originating, e.g., from not modelled break lines.

For the interpolation we use the “moving plane” method. The plane is stopping at the break line, and will not eliminate the points beyond it. (see figure below).

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.

On the right is the DHM derived by the linear prediction without any manual intervention.

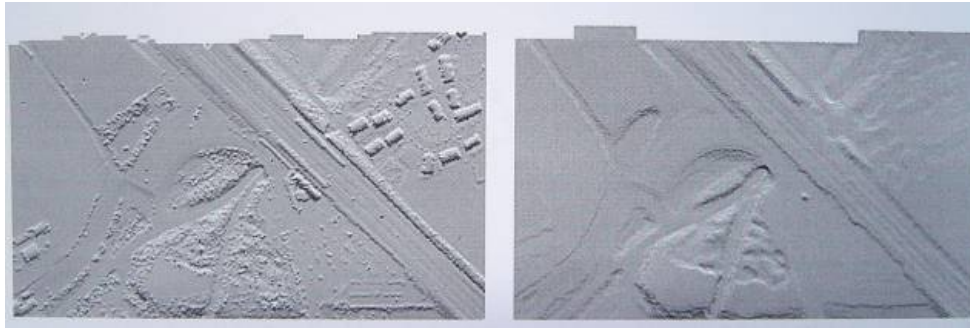


Fig. 3 Surface interpolated using the original data.
DHM derived by the filtered data.

2.2. Program Rascor for filtering and interpolation

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.

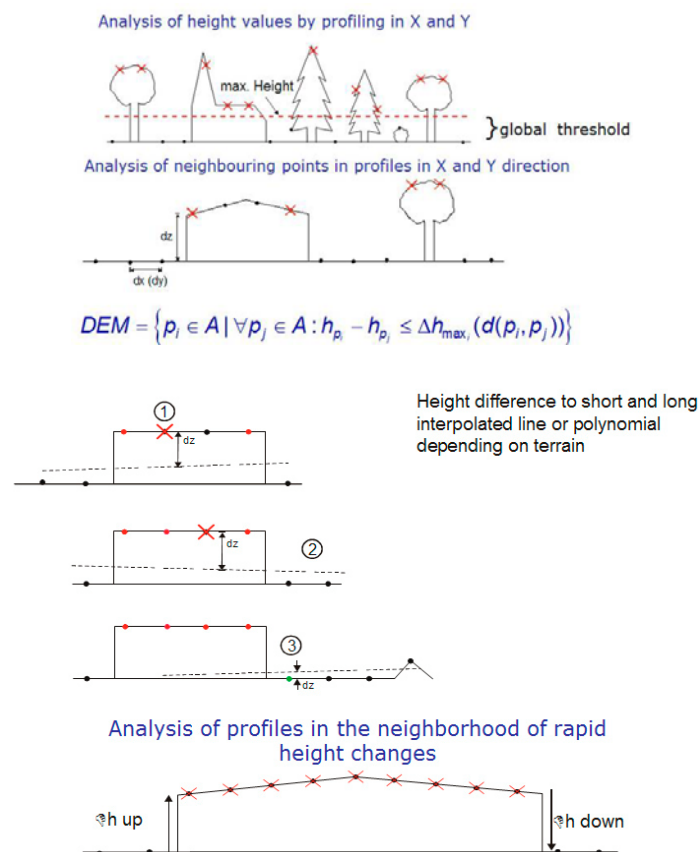


Fig. 4 Different terrain analysis.

The interpolated surface in the point P_i is given by:

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

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

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

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

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

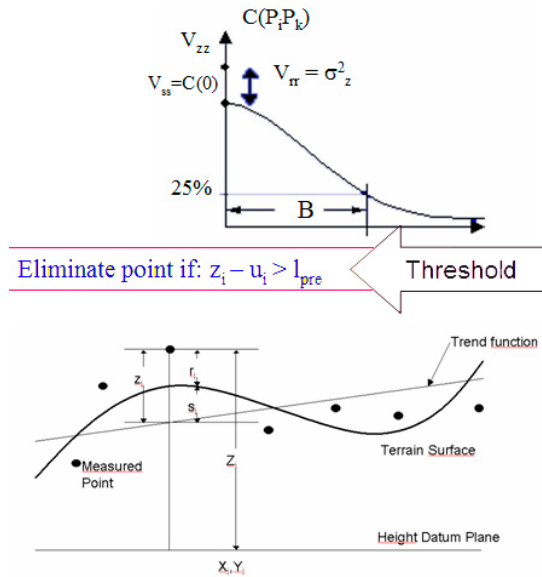


Fig. 5 Interpolation method

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 semi-automatically acquired and needs a supervisor. They also can be manually imputed or added in the iterative filtering process.

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

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

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 derivates of the Gauss-function:

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

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

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

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

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

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

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.

Highway

Highways are very often located on dams or in cuts, so it corresponds to railways and dams. A general filtering without taking care about the special classes tends to an elimination or reduction of the dams, especially the upper shoulders. If only these special segments are filtered, the filter parameters will take care about the special structure but nevertheless the upper dam shoulders are influenced.

Bridges

Bridges do show a sharp change of the height without any slope. Of course, if special segments are used just for the bridges, they will not be changed because points are only located on top of the bridges without more or less variation in height. The break lines can include also the neighboured dams, so a handling may be easier.

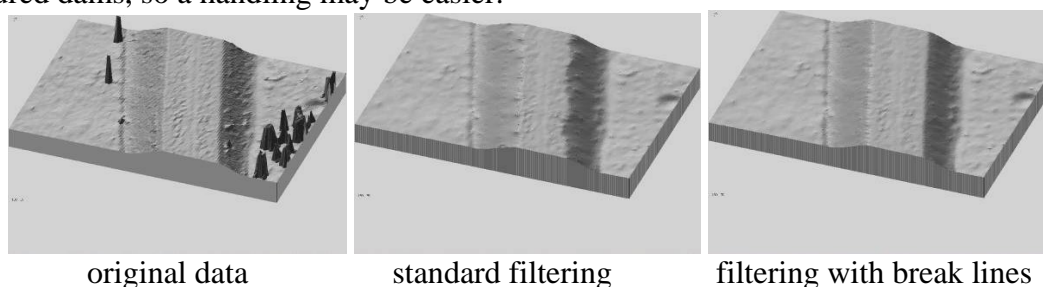


Fig. 6 Example of a highway

These elements have to be handled in a special manner. RASCOR allows the definition of such special areas where only a local filter will be used for the elimination of points located on cars and on bushes on the slope. Another possibility is the use of the segment limits and also the roads as break lines. The program will not take care about points located behind break lines.

As it can be seen in figure 14, a standard filtering of a highway dam is removing the trees and also bushes on the slope, but it takes also points out of the shoulder. A standard filtering removed 21.6% of the points. If break lines are taken into account, only 10.4% of the points were excluded and the dam looks quite more like it should be.

Slopes

Slopes do show similar problems like highways and dams. They can be handled as special areas, but it is better to use break lines instead of this.

Caused by the mirror effect of the water surface, usually there is no signal available for water bodies, resulting in missing height information about the water surface. But just beside the water, height information is available which is sufficient for the height definition of the water surface. Typical characteristics of watercourses are slopes and also parallel dams. Cuts are not causing problems because they are not changed by RASCOR, but problems exist with dams beside the watercourse if they have not been separated.

A global filtering of the whole data set is affecting the dam parallel to the water body while the individual filtering is not causing problems. The effect of separate handling of the water bodies is achieved by the introduction of break lines.

Grass

Meadows and willows are simple segments; they do include only few points not belonging to the bare surface like located on bushes and trees, so they can be handled with the global and the segmented filtering without any problems.

Vegetation

As vegetation areas a mixture of grass and bushes are defined. These segments could be handled in a similar manner like the segment grass. Corresponding to this, the segmented as well as the global filtering yielded to very similar results.

Forest

A global filtering of the whole area was done only with one iteration because a second iteration affected especially dams and highways. For the forest area a second iteration is recommended. The filter parameters are not so sensitive. With a second iteration also the global filtering was leading to sufficient results in the forest area, but as mentioned, other areas have been influenced.

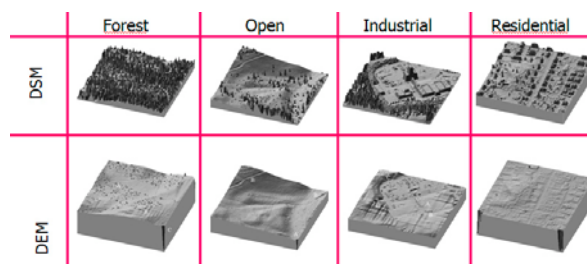


Fig. 7 Results in different type of areas

2.3. Results with RASCOR

Example 1: study area : Williamstown USA.

RASCOR can be handled in the batch mode or a sequence of batch modes. The batch-handling is based on the file rascor.dat which may include a switch for the automatic handling as batch job.

2.3.1. Using break lines

ANALYSIS AND FILTERING OF DIGITAL TERRAIN MODELS

FILE NAMES

CODE	FILE / FUNCTION	NAME
1	INPUT FILE	grid.dat
2	INPUT FILE WITH POINT NAMES ? Y/N	N
3	OUTPUT FILE OF FILTERED DEM DATA	daxyz.dat
4	FILE FOR PLOT DATA	
5	FILE FOR EXCLUSION	
8	FILE FOR OUTPUT OF REMOVED POINTS	
9	FILE WITH BREAK LINES	williamstown-nj.brk
10	BREAK LINES WITH 1 CODE (1) 2 CODES (2)	
	POINT NAMES (P) NO POINT NAMES (N)	2

FILE NAME = BLANK = NO CREATION OF FILE

TYPE CODE AND FILE NAME IN ONE LINE DEFAULT = NEXT INPUT

OPTIONS 1

- 1 LISTING INPUT VALUES ? Y/N N
- 2 LISTING SHORT (S) OR LONG (L) S/L/V S
- 3 INTERPOLATION(I) OR ELIMINATION (E) E
- 4 OUTPUT WITH POINT NAMES? N/S/R N=NO
S=NAMES=SEQUENCE R=NAMES=RASTER S
- 5 SAME TYPE OF TERRAIN(S) VARYING (V) S
- 6 FLAT (F), ROLLING (R)
OR MOUNTAINOUS (M) F/R/M R
- 7 NUMBER OF ITERATIONS 1 / 2 2
- 8 USE MODE CHANGE UP/DOWN Y/N Y
- 9 SMOOTHING AFTER FILTERING 0 = NO
FINAL FILTER - SIZE 0/3/5/7/9 0
- 11 OUTPUT FILE FOR LISA Y/N N
- 12 FURTHER HANDLING D = DIALOG
A=AUTOMATIC = NO FURTHER DIALOG D/A
PREDICTION YES/NO DY/DN/AY/AN AN
- 13 SIMPLE / ENHANCED FILE INPUT S/E S

FILE NAME = BLANK = NO CREATION OF FILE

TYPE CODE AND FILE NAME IN ONE LINE DEFAULT = NEXT INPUT

Final results:

384523 OUTPUT POINTS (65.10 %) INCLUDING SPECIAL POINTS

206165 POINTS NOT USED (34.90 %)

END OF PROGRAM RASCOR

2.3.2. Without consider break lines:

Final results:

301908 POINTS NOT USED (51.11 %)

Program RASCOR creates an image situ.raw with the location of the remaining points. If a break line file will be used, situ.raw will include an overlay of the break lines (gray value 0) and the remaining points (gray value 120).

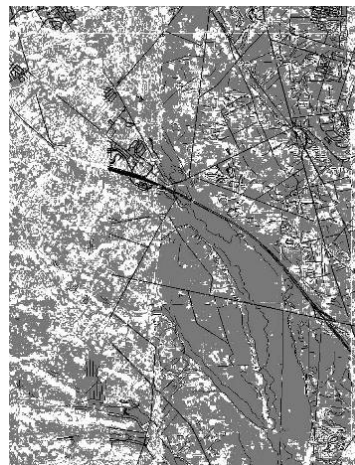


Fig.8 (left) file situ.raw – remaining points, without break lines, 51,11% of points removed
(right) file situ.raw – remaining points, with break lines, 34,90% of points removed

The break lines can be modified by operator .

The file wich contains the coordinates for the break lines can be modified with a text editor.

The following example shows the structure of this file. Break lines are coded with 0 and 1 values which represent the start and the end of a break line.

```
#INDEX HEADER    X      Y      Z
.....
2610 1    352073.824 331166.299 171.781
added break lines start here.
3000 0    356751.300 319548.010
3001 1    356907.460 319552.770
3002 1    357891.270 318959.360
3016 0    361888.980 317491.460
3007 1    357141.700 319740.170
```

The following images shows the influence of the added breaklines.



Fig.9 Fewer points are removed 34,82%.

2.4. Visualisation of results

LISA-BASIC is a general DEM-program which has been used to represent the results obtained with Rascor. Based on random or equal distributed height points, supported by break and form lines an equal distributed DEM can be created. Internally the DEM is handled as an image with grey- instead of Z-values. This has the advantage of a very fast computation of all derived products like contour lines, wire frame models or 3-D-views.

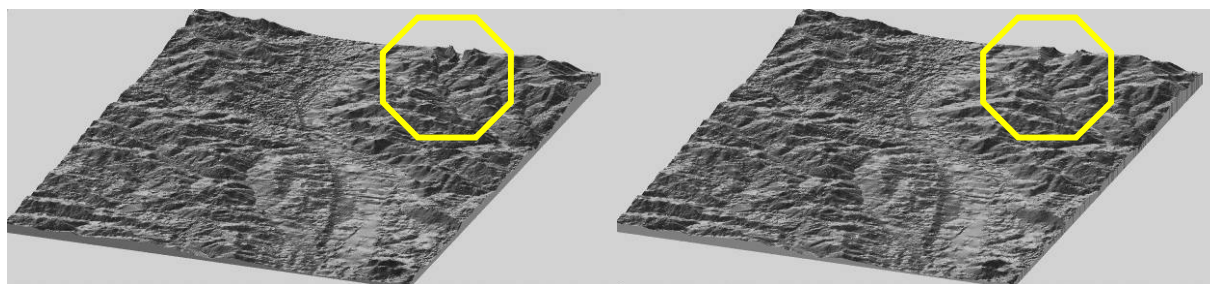


Fig.10 Block.ima with break lines – exaggeration 25%

Fig.11 Block.ima without break lines – exaggeration 25%

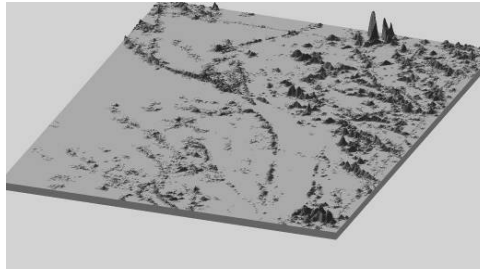
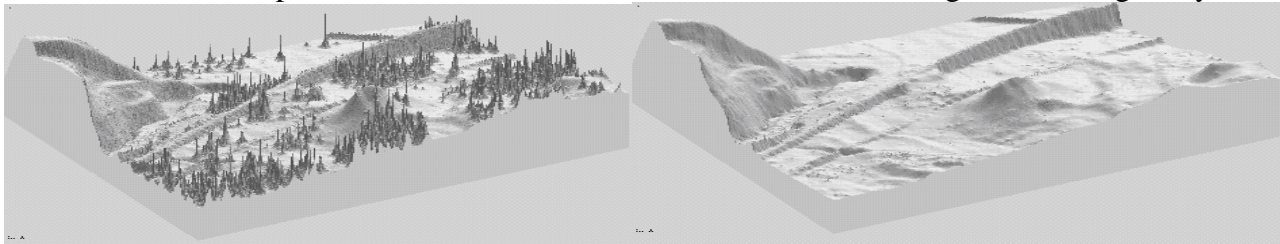


Figure 12 Differential DHM (clarifies in a graphical representation all the differences between the two models)

Another example is a undulated ared which contains trees and a segment of a high-way.



Original data set

After refinement with RASCOR

Figure 13

Example 2: study area : A pipe crossing a valley – Romania.

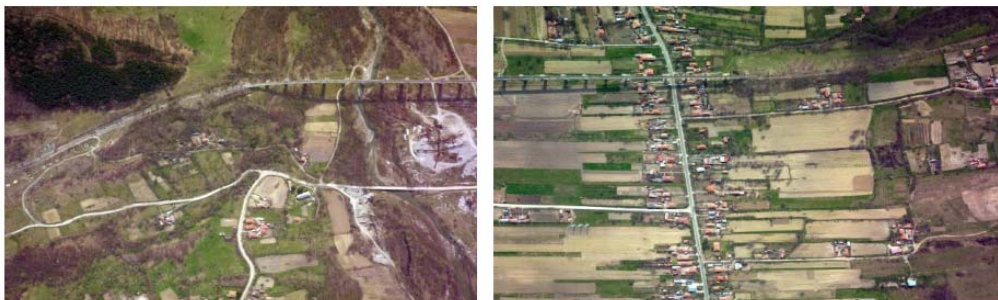
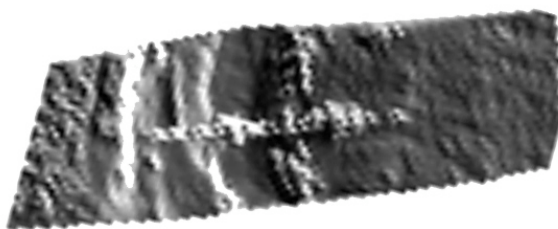
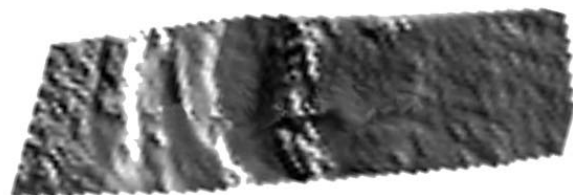


Figure 14 Aerial photos for the area of interest



before filtering



after refinement with Rascor (without break lines)

Figure 15

3. Conclusions

Automatic filtering of a DSM \rightarrow 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 example showed in this paper-work is suggestive.

The programs used (RASCOR and Lisa developed at IPI) are very good . The filtering and interpolation works very fast and in a simple to use interface , but more important : the program RASCOR is using now break lines for filtering , which is a big improvement because the results are more close to the real situation of terrain.

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.

Bibliography

1. 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"
2. Jacobsen, K. 2003, *RASCOR – manual*, Institute for Photogrammetry and Engineering Survey ,University of Hannover, Germany
3. 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
4. Passini R., Betzner D., Jacobsen K., 2002, *Filtering of Digital Elevations Models*, ASPRS annual convention, Washington 2002
5. K. Kraus and N. Pfeifer 2002, *ADVANCED DTM GENERATION FROM LIDAR DATA*
6. *Institute of Photogrammetry and Remote Sensing Vienna University of Technology*
7. 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
8. 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*
9. Krysia Sapeta, *LIDAR, Analytical Surveys Inc. (ASI), Colorado Springs, Colorado*

Aknoledgements:

Special thanks to Institute for Photogrammetry and GeoInformation Hannover for support.