SLOPE BASED FILTERING OF LASER ALTIMETRY DATA

George VOSSELMAN
Department of Geodesy, Faculty of Civil Engineering and Geosciences
Delft University of Technology, The Netherlands
g.vosselman@geo.tudelft.nl

Working Group III/2

KEY WORDS: Laser, DEM, Filtering

ABSTRACT

Laser altimetry is becoming the prime method for large scale acquisition of height data. Although laser altimetry is full integrated into processes for the production of digital elevation models in different countries, the derivation of DEM's from the raw laser altimetry measurements still causes problems. In particular the laser pulses reflected on the ground surface need to be distinguished from those reflecting on buildings and vegetation. In this paper a new method is proposed for filtering laser data. This method is closely related to the erosion operator used for mathematical grey scale morphology. Based on height differences in a representative training dataset, filter functions are derived that either preserve important terrain characteristics or minimise the number of classification errors. In experiments it is shown that the latter filter causes smaller errors in the resulting digital elevation models. In general the performance of the filters deteriorates with a decreasing point density.

1 INTRODUCTION

Laser altimetry is becoming the prime method for large scale acquisition of height data. Several countries are currently using laser altimetry for creating or updating very detailed regional or nation-wide digital elevation models. In the Netherlands, for example, a nation-wide digital elevation model is being created with a density of 1 point per 16 m2 [Huising and Gomes Pereira, 1998]. Even though production lines are set up, the derivation of digital elevation models from the measurements by airborne laser scanners is not without problems. The two major problems are the elimination of systematic errors and the selection of ground points [Huising and Gomes Pereira, 1998].

The presence of errors in laser altimetry data often becomes evident when combining data from adjacent strips. In the overlap between the strips systematic differences between the heights of the points and the location of height jumps can be observed. A topic of research is to model these errors and to eliminate them in a so-called strip adjustment, much like eliminating lens distortion in a self-calibrating block adjustment [Kilian et al., 1996].

The other problem is the selection of the ground points. The pulses emitted by the laser scanner can be reflected by buildings, trees, cars, electricity wires, and many other objects on top of or above the ground surface. The height of these objects should not be included in the conventional digital elevation models. Therefore, filter methods are used to discriminate the ground points from other points. In this paper, a new approach to filtering is presented.

Lindenberger [1993] describes how mathematical morphology can be used for filtering data recorded by a laser profiler. A first estimation for the ground surface is obtained by an opening on the recorded data with a horizontal structure element. All points within some distance of the estimated ground surface are classified as ground points. An auto-regressive process is used to improve the results obtained by the opening. The application of the auto-regressive process requires an ordering of the laser points. Therefore it is suited for processing data obtained by laser profilers. Data obtained by laser scanners is, however, scattered in the XY-plane, for which there is no logical one-dimensional ordering. In contrast to the auto-regressive process, the opening and selection of points within some distance of the resulting surface can easily be extended to two-dimensionally scattered data.

Kilian et al. [1996] note that the size of the structure element used for the opening is a critical parameter for which there is no single optimal value. They suggest to use a series of openings with different structure element sizes. For each point, the maximum size at which this point is within some distance of the opened surface is assigned as a weight to this point. These weights are used in a final smoothing step to estimate the ground surface.

Pfeifer et al. [1998] describe a filter method based on an iterative linear least squares interpolation. Using a weight function which assigns low weights to relatively high points, a robust estimation of the ground surface is obtained. The weight function is described by four parameters.

All methods described above use the assumption of a locally horizontal terrain or a uniform point distribution to determine the ground points. In order to preserve the ground points in sloped terrain, the window sizes need to be restricted or distance thresholds need to be increased. The optimal values of the filter parameters clearly differ from one terrain type to another. Based on experience with these filters, a list of optimal parameter values can be made, but there is no direct relationship between these values and characteristics of the terrain.

In this paper a new filter method is presented in which the height differences between ground points are used to determine the optimal filter function. The next section describes the principle of this method, the relationship to mathematical morphology, and some implementation aspects. In section three several methods for deriving a filter function are introduced. Different filter objectives will lead to different filter functions. In section four and five the set-up and results of a series of experiments are presented. Conclusions are given in the last section.

2 SLOPE BASED FILTERING

2.1 Filtering principle

The basic idea, like in the methods mentioned above, is based on the observation that a large height difference between two nearby points is unlikely to be caused by a steep slope in the terrain. More likely, the higher point is not a ground point. Clearly, for some height difference, the probability that the higher point could be a ground point decreases if the distance between the two points decreases. Therefore, Kilian et al. [1996] introduce weights depending on the size of the morphological filter kernel and Pfeifer et al. [1998] implicitly weigh the heights by a covariance function which depends on the distance between two points.

Instead of introducing weights, we explicitly define the acceptable height difference between two points as a function of the distance between the points: $\Delta h_{\max}(d)$. In general, this will be a non-decreasing function. In section three, several methods for deriving such a function will be described.

The filter function can now be used to define the set of points that are classified as ground points. Let A be the set of all points and DEM be the set of ground points, then

$$DEM = \left\{p_i \in A \mid \forall p_j \in A : h_{p_i} - h_{p_j} \leq \Delta h_{\max}\left(d(p_i, p_j)\right)\right\} \tag{1}$$

In words: a point p_i is classified as a terrain point if there is no other point p_j such that the height difference between these points is larger than the allowed maximum height difference at the distance between these points. This filter definition is closely related to some concepts from (grey scale) mathematical morphology.

2.2 Relation to mathematical morphology

The erosion $e(x,y)$ of a two-dimensional signal $h(x,y)$ with a kernel $k(\Delta x,\Delta y)$ is defined as [Haralick and Shapiro, 1992]:

$$e(x,y) = \min_{\Delta x} \min_{\Delta y} \left[h(x+\Delta x, y+\Delta y) - k(\Delta x, \Delta y) \right] \tag{2}$$

For a point p_i of a discrete set of points A, this corresponds to

$$e_{p_i} = \min_{p_j \in A} \left[h_{p_j} - k(x_{p_j} - x_{p_i}, y_{p_j} - y_{p_i}) \right] \tag{3}$$

If we define the kernel function as

$$k(\Delta x, \Delta y) = -\Delta h_{\max}\left(\sqrt{\Delta x^2 + \Delta y^2}\right) \tag{4}$$

(note the minus sign!), the eroded value at point p_i can be written as

$$e_{p_i} = \min_{p_j \in A} \left[h_{p_j} + \Delta h_{\max} \left(d(p_i, p_j) \right) \right] \tag{5}$$

Now the relationship between the erosion and the defined ground point filter can be expressed. If $h_{p_i} \le e_{p_i}$, then

$$\forall p_j \in A : h_{p_i} \le h_{p_j} + \Delta h_{\max} \left(d(p_i, p_j) \right) \tag{6}$$

Thus, the set of ground points can be defined by

$$DEM = \left\{ p_i \in A \mid h_{p_i} \le e_{p_i} \right\} \tag{7}$$

In words: a point is classified as a ground point if its height does not exceed the height of the eroded surface. Filtering laser altimetry data can therefore be done with two elementary operations: a morphological erosion with the kernel function as defined in (4) and a test on the difference between the original height and the eroded height of a point.

2.3 Implementation issues

Morphological operators are available in most image processing packages. These packages usually require the data to be organised in a grid. Interpolating the raw laser altimetry data to a grid, however, causes a significant loss of information. In particular when heights are interpolated between ground points and points on vegetation or buildings, the height differences in the interpolated data will be reduced. Therefore, it becomes more difficult to make a correct classification. Although it is computationally more expensive, we prefer to work with the original, irregularly distributed point data.

According to the filter definition in equation (1), the height of a point needs to be compared with the heights of all other points. In most cases this is not necessary. For example, if it is known that the height difference in the terrain is no more than 10 m and that $\Delta h_{\max}(100\text{ m}) = 10$ m, only points within 100 m of the current point need to be considered. In order to determine the points within some distance, we organise the points in a Delaunay triangulation.

Instead of verifying that a point fulfils the condition in equation (6), one can also check whether the height of a point results into the rejection of a point in its neighbourhood. This approach has been taken in the implementation. First the height of the direct neighbours in the triangulation are verified. Similar to a region growing algorithm, in the next step the neighbours of these neighbours are verified. The neighbourhood is grown until there are no more adjacent points within the maximum distance that needs to be considered. To speed up the filtering process, the neighbours of a node are only verified if the node itself is rejected. This implies that points in the DEM set may not strictly comply with the definition in equation (7). The number of errors introduced by this heuristic is very small. The processing time, however, reduces with a very large factor (10-100).

3 DERIVATION OF FILTER KERNELS

The filter function should incorporate the knowledge about the height differences in the terrain. In this section we describe three ways to derive a filter function. The first one uses simple generic knowledge about the terrain shape and the precision of the height measurements. The other two make use of a training data set.

3.1 Synthetic function

Suppose that, for some area, one would know that the slopes in the terrain are not steeper than, say, 30%. If the measurements are free of errors, the filter function could be defined as:

$$\Delta h_{\max}(d) = 0.3\, d \tag{8}$$

In the usual case of noisy measurements, one would like to add a confidence interval. If one would allow that 5% of the terrain points with a standard deviation σ may be rejected, the filter function becomes

$$\Delta h_{\max}(d) = 0.3\, d + 1.65\sqrt{2}\sigma \tag{9}$$

3.2 Preserving important terrain features

In most cases it will be difficult to specify a filter function in terms of parameters (as above). An alternative is to derive the shape characteristics of the terrain from a training sample. This part of the data should contain the important terrain features that should be preserved by the filter and should only consist of ground points. If one can specify such an area, the points in this area can be used to empirically derive the maximum height differences as a function of the distance between two points.

Clearly, the determined maximum height differences are stochastic. Before using these values to filter the points in other areas, one should again add a confidence interval for the maxima. Suppose the training sample contains N point pairs at a distance d. The maximum height difference at this distance can be considered to be the maximum of N drawings out of the height differences of the complete dataset. Unfortunately, the probability distribution of height differences in the complete dataset is not known. To get some idea about the standard deviation of the maximum height difference, we therefore make use of the known distribution of height differences in the training sample.

Let $\mathrm{F}(\Delta h)$ be the cumulative probability distribution for the height differences between points at distance d in the training data. Then, $\mathrm{F}_{\max}(\Delta h) = \mathrm{F}(\Delta h)^N$ is the cumulative probability distribution for the maximum of N independent drawings of a height difference. The probability density function of the maximum becomes

$$\mathrm{f}_{\max}(\Delta h) = \frac{\partial\, \mathrm{F}_{\max}(\Delta h)}{\partial\, \Delta h} = N\, \mathrm{F}(\Delta h)^{N-1}\, \frac{\partial\, \mathrm{F}(\Delta h)}{\partial\, \Delta h} = N\, \mathrm{F}(\Delta h)^{N-1}\, \mathrm{f}(\Delta h) \tag{10}$$

The variance of the maximum can easily be obtained by taking integrals over this function. For each distance d the variance needs to be computed independently. From the variance a confidence interval can be derived that should be added to the maximum height difference that was encountered for distance d.

3.3 Minimising classification errors

Whereas the above filter function makes sure that important terrain features are maintained in the digital elevation model, it may be quite liberal in accepting points that are not on the ground. I.e., the number of type I errors will be low, but the number of type II errors may become large.

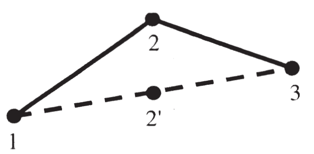

Figure 1. DEM error.

Suppose, a dataset contains three points on a line, as in figure 1, and that one would use linear interpolation between points to determine the heights at other points. If point 2 would be rejected, but would in fact be a ground point (type I error), the height error in the resulting DEM at the position of point 2 is h_2 - $h_{2'}$. If, on the other hand, point 2 would be accepted as a ground point, but would in fact be a point on a building (type II error), the height error in the DEM at the position of point 2 is $h_{2'}$-h_2. Hence, the absolute size of a height error caused by a wrong classification is the same for type I and type II errors. Since the effect of these errors is the same, a point p_i can best be classified as a ground point if $P(p_i \in DEM) > P(p_i \notin DEM)$. The break even point lies at $P(p_i \in DEM) = 0.5$. Knowing the height of a ground point p_j at distance d from point p_i, one would like to determine the height difference Δh between the points, such that $P(p_i \in DEM \mid \Delta h, d, p_j \in DEM) = 0.5$. Using probabilities derived from frequency counts of height differences between point pairs in a training set of ground points and between point pairs with one point from the training set of ground points and the other point from the training set of unfiltered data of the same area, one can calculate

$$\begin{aligned} P(p_i \in DEM \mid \Delta h, d, p_j \in DEM) &= \frac{P(p_i \in DEM, \Delta h, d, p_j \in DEM)}{P(\Delta h, d, p_j \in DEM)} \\ &= \frac{P(\Delta h \mid d, p_i \in DEM, p_j \in DEM)\;\; P(p_i \in DEM \mid d, p_j \in DEM)}{P(\Delta h \mid d, p_j \in DEM)} \end{aligned} \tag{11}$$

for each height difference Δh and distance d. Both Δh and d need to be discrete for this purpose. For each d, one can now determine the Δh for which $P(p_i \in DEM \mid \Delta h, d, p_j \in DEM) = 0.5$. These values Δh can be taken as the maximum height differences that are allowed in the filtered data in order to minimise the number of classification errors.

4 EXPERIMENTS

Several experiments have been performed to obtain insight into the behaviour of the developed filter method. The next paragraphs discuss the used dataset, the error measures, and the derived filter functions.

4.1 IJsselstreek data set

The data used in this test has been recorded by the helicopter based FLI-MAP system [Huising and Gomes Peirera, 1998, Baltsavias, 1999]. The point density varies between 5 and 7 points per m^2. The FLI-MAP system is primarily designed for monitoring roads, rail-roads and power-lines and, therefore, only records the first returning laser pulse. Consequently, the penetration rate in areas with vegetation is relatively low. A small part with high vegetation, a meadow, and a dike with about 1.3 million points was selected (figure 3, upper left image). The average point density in this area was 5.6 points per m^2. Roughly half the points are ground points.

4.2 Ground "truth"

Unfortunately, it is virtually impossible to establish accurate ground truth for laser altimetry data. This holds in particular for areas with dense vegetation. Ground truth for the test area was not available. For the analysis of the filter results, we made use of the filter property that the classification results improve with the point density. E.g. if there is a height difference of 1 m between two points at a distance of 0.5 m, the higher point is most likely not a ground point. If, however, the point distance is 4 m, a height difference of 1 m could as well be explained by variations in the terrain height. The filter results obtained with the original dataset were therefore used as a reference for the filter results of datasets with a reduced point density. From the original dataset, datasets with average point densities of one point per 1, 4, and 16 m^2 were derived.

4.3 Derived filter functions

A training set with a part of a dike and some vegetation was selected for the derivation of the filter functions described in section 3.2 and 3.3. These filter functions will be called the maximum filter and the probabilistic filter. The training set contained 43000 points from which about 69 million height differences between points within a distance of 10 m could be computed. The resulting functions are shown in figure 2. It was found that the standard deviations of the maximum height differences were quite small, usually below 1 cm. Thus only a small confidence interval was added to the maxima found in the training set. Figure 2 clearly shows that the maximum filter allows larger height differences than the probabilistic filter.

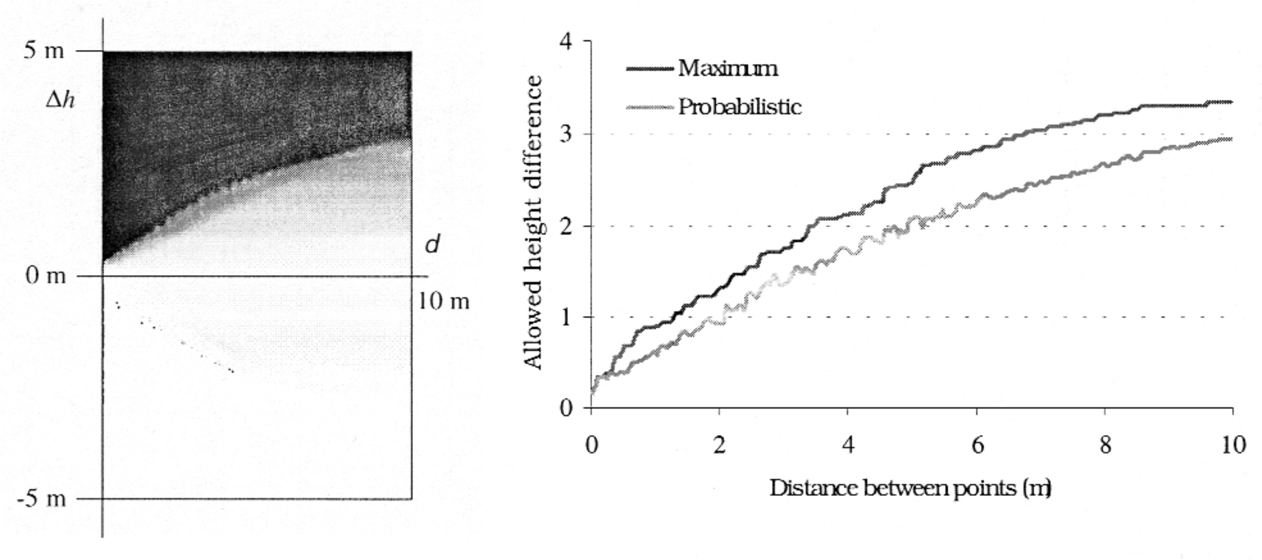

Figure 2. Filter functions derived from training data. Left: Probabilities $P(p_i \in DEM \mid \Delta h, d, p_j \in DEM)$. Black is 0.0, white is 1.0. Right: The upper curve shows the maximum height differences between ground points. The lower curve shows the height differences for which $P(p_i \in DEM \mid \Delta h, d, p_j \in DEM) = 0.5$.

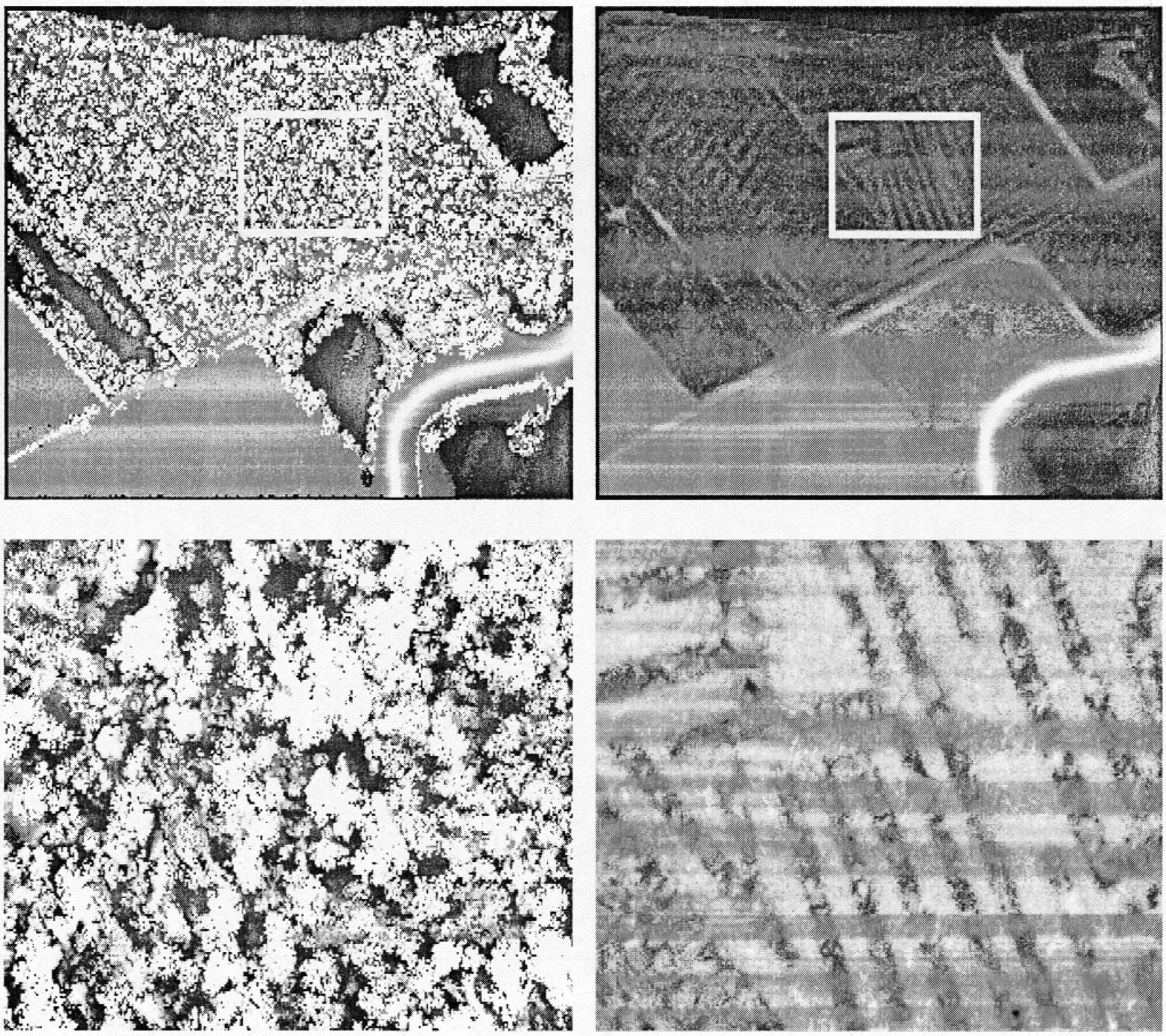

Figure 3. Unfiltered (left) and filtered (right) laser data with a point density of 5.6 points/m2 shown in a grid of 2 m (top) and 0.5 m (bottom). The filter were results obtained with the maximum filter. All points higher than 8 m are imaged white. The contrast in the lower right image is stretched. The height difference at the ditches in this image is only 0.5 m.

5 RESULTS

After filtering the original data with the maximum filter, the resulting DEM showed a detailed structure of ditches below the vegetation (figure 3). In the left upper image, some vegetation points can be found just above the dike. This area was under water. Since there are no measurements on the water surface, the filter selected the lowest points of the vegetation. In the enlarged part, small errors due to low vegetation can be observed. The brighter spots inside the ditches most likely are points on shrubs. Thus, it has to be kept in mind that these filter results, that are used as a reference for the results on datasets with a lower point density, are certainly not perfect.

The percentages of type I and II errors in the results of filtering the datasets with a reduced resolution are given in table 1. As to be expected, the number of type I errors is smaller for the maximum filter than for the probabilistic filter. Since the maximum filter allows larger height differences, the number

Point density	Type I		Type II		Total	
(points/m^2)	max	prob	max	prob	max	prob
1	0.0	2.3	5.0	2.0	2.3	2.2
1/4	0.0	1.1	8.7	5.2	4.0	3.0
1/16	1.8	3.1	13.9	10.1	7.3	6.3

Table 1. Percentages of type I and II errors

of type II errors for this filter is larger than for the probabilistic filter. For both filters, the number of errors, in particular the number of type II errors, becomes larger for the datasets with a lower point density. In the probabilistic filter the maximum height differences were derived such that they minimise the number of errors. In table 1 the total amount of errors indeed appears to be smaller in case of the probabilistic filter.

Point density (points/m^2)	Mean error		RMS error		Max. error	
	max	prob	max	prob	max	prob
1	0.01	0.00	0.09	0.07	2.72	1.97
1/4	0.03	0.01	0.17	0.11	3.05	3.05
1/16	0.06	0.04	0.30	0.22	4.22	3.20

Table 2. Error size statistics (m)

More important than the number of errors is the effect of these errors onto the digital elevation model. For this purpose the heights of type II errors were interpolated in the TIN of the filtered reference points and the heights of the type I errors were interpolated in the TIN of the filter results of the reduced dataset. The statistics on the errors sizes are shown in table 2. Again the results of the probabilistic filter are better than the results of the maximum filter. Most classification errors are relatively small. Therefore, the mean error remains quite small. The root mean square error values clearly increase if the point density decreases. This is also illustrated in figure 4. In the DEM reconstructed from one point per 16 m2 several height variations can be seen that are caused by unfiltered points in vegetation. Also the ditches (about 5 m wide and 50 cm deep) can not longer be reconstructed.

The maximum errors in this test became quite large, sometimes even larger than the maximum value in the filter functions. This is caused by points that had no other point within a distance of 10 m. Since this was the maximum range of the derived filter function, these kind of points could not be filtered.

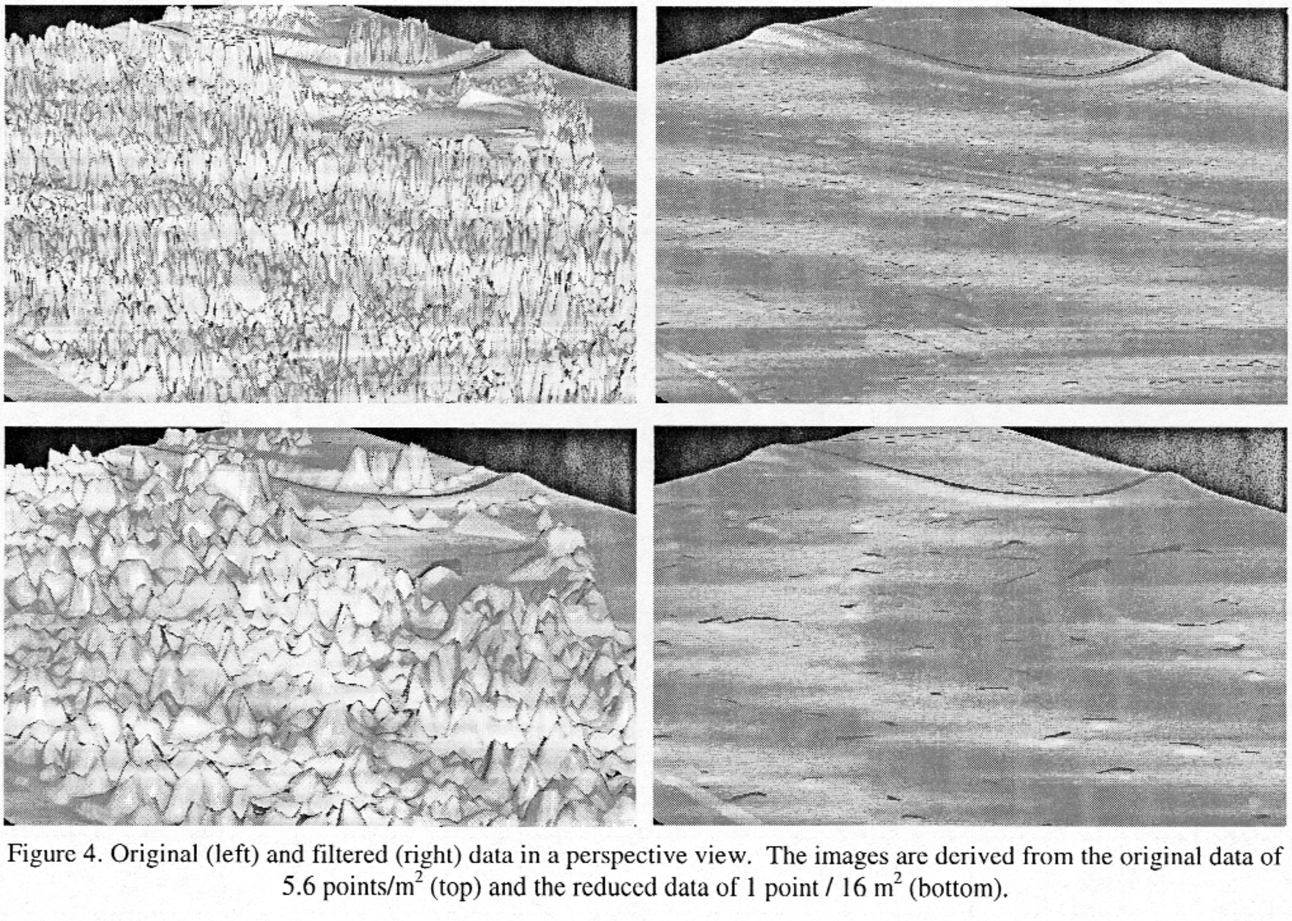

Figure 4. Original (left) and filtered (right) data in a perspective view. The images are derived from the original data of 5.6 points/m^2 (top) and the reduced data of 1 point / 16 m^2 (bottom).

6 CONCLUSIONS

In this paper a method has been presented for filtering laser altimetry data. The method is closely related to the erosion operator used in mathematical morphology. The shape of the filter function can be derived from a set of training data. It was shown that a function which minimises the probability of a classification error produces a DEM with smaller errors than a function which tries to preserve shape characteristics in the training data.

As expected, the filter results deteriorate with an decreasing point density. The filtering of reflections on low vegetation can not be perfect, and will always cause errors in the derived digital elevation model. Although the mean errors in the computed DEM's were relatively small (4-6 cm for 1 point per 16 m^2), the RMS error values were found to be in the order of 20-30 cm. The true precision of the DEM is expected to be slightly worse, since the effects of the measurement noise was not included in the computed RMS values.

In this paper points were classified solely by comparing height differences between two points. Better classifications can be expected if other features, like height textures derived from multiple points are also used [Maas, 1999, Oude Elberink and Maas, 2000]. In that case it should become easier to make a distinction between a ground point on a sloped surface and a vegetation point on a horizontal surface, even though the maximum height differences between these points and their surrounding points are the same.

Another way to improve the classification is to introduce support from an image analyst. If the morphological characteristics of a terrain vary within an area to be processed, one could (by roughly drawing polygons on an image) indicate areas which are more or less homogeneous. For each type of terrain one could then use a different training set to derive the optimal filter for that terrain type. The amount of required interaction is quite low, but the filter results could improve considerably.

ACKNOWLEDGMENTS

The laser altimetry data was kindly provided by the Survey Department of the Ministry of Transport, Public Works and Water Management.

REFERENCES

Baltsavias, E.P., 1999. Airborne laser scanning: existing systems and firms and other resources. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 54, No. 2-3, pp. 164-198.

Haralick, R.M., Shapiro, L.G., 1992. Computer and robot vision. Vol. 1, Addison-Wesley, Reading, MA.

Huising, E.J., Gomes Pereira, L.M., 1998. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 53, No. 5, pp. 245-261.

Kilian, J., Haala, N., Englich, M., 1996. Capture and evaluation of airborne laser scanner data. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, Vienna, pp. 383-388.

Lindenberger, J., 1993. Laser-Profilmessungen zur topographischen Geländeaufnahme. Deutsche Geodätische Kommission, Series C, No. 400, Munich.

Maas, H.-G., 1999. The potential of height texture measures for the segmentation of airborne laser scanner data. Fourth International Airborne Remote Sensing Conference and Exhibition / 21st Canadian Symposium on Remote Sensing, Ottawa, Ontario, Canada, 21-24 June 1999, pp. I-154-161.

Pfeifer, N. Köstli, A., Kraus, K., 1998. Interpolation and filtering of laser scanner data - implementation and first results. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXII, Part 3/1, Columbus, pp 153-159.

Oude Elberink, S., Maas, H-G., 2000. The use of anisotropic height texture measures for the segmentation of airborne laserscanner data. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Commission III, Amsterdam.