Advances in Geosciences Konstantinos Perakis, Editor
EARSeL, 2012

Semi-automatic detection of burial mounds in forested areas

Øivind Due Trier[1], Maciel Zortea[1] and Siri Øyen Larsen[1]
[1]*Norwegian Computing Center, Section for Earth Observation, P.O. Box 114 Blindern, NO-0314 Oslo, Norway; oivind.due.trier@nr.no, maciel.zortea@nr.no, siri.oyen.larsen@nr.no*

Abstract. This paper describes a new method for the automatic detection of heap structures in airborne laser scanning data, and reports on a work in progress. The heaps could be ancient grave mounds, dating to 1500-2000 years ago. Such grave mounds are automatically protected in Norway. Several Norwegian municipalities are experiencing growing pressure on forested land for development, being it new residential areas, industry, tourism, or new highways. The traditional mapping of cultural heritage, mainly based on chance discovery and inaccurate positioning, has proven inadequate for land use planning. Therefore, the Norwegian Directorate for Cultural Heritage, in cooperation with some counties and municipalities, are investing in the development of new methods, using new technology, for a more systematic mapping of cultural heritage.

Grave mounds are one of the most frequent types of archaeological structure in Norway. We have earlier developed a method for the automatic detection of circular soil marks and crop marks in cereal fields in satellite and aerial images. Several of these detections have been confirmed to be leveled grave mounds.

Methods based on optical images are of limited value in forested areas, since the archaeology tends to be obscured by the tree canopies. However, by using lidar data, the forest vegetation can be removed from the data, making it possible to detect archaeology in a semi-automatic fashion, provided the archaeology manifests itself as structures in the digital elevation model of the lidar ground returns, and that these structures may be described using some kind of pattern. In the majority of Norway's 19 counties, there are intact grave mounds in forested areas. This means that a semi-automatic method for the detection of grave mounds in lidar data would be an important tool in a more systematic mapping of archaeology in Norway.

The automatic method has been applied on lidar data from Larvik municipality in Vestfold County, Norway. Preliminary results are promising, and indicate that this may be a very useful tool for archaeologist in Norway for a more systematic mapping of cultural heritage.

Keywords. Grave mounds, airborne laser scanning, lidar, pattern recognition, confidence estimation.

1. Introduction

Several Norwegian municipalities are experiencing growing pressure on forested land for development, being it new residential areas, industry, tourism, or new highways. The traditional mapping of cultural heritage, mainly based on chance discovery and inaccurate positioning, has proven inadequate for land use planning. Therefore, the Norwegian Directorate for Cultural Heritage, in cooperation with some counties and municipalities, are investing in the development of new methods, using new technology, for a more systematic mapping of cultural heritage.

One of the most frequent types of archaeological structure in Norway is grave mounds (Figure 1). We have earlier developed a method for the automatic detection of circular soil marks and crop marks in cereal fields in satellite and aerial images [6]. Several of these detections have been confirmed to be levelled grave mounds, dating to 1500-2500 years ago.

Figure 1. Examples of grave mounds, Larvik municipality, Vestfold County, Norway. Top: a grave mound in Bøkeskogen, with a thin layer of snow. Bottom: a grave mound in Brunlafeltet, with a looting pit in the middle.

Methods based on optical images are of limited value in forested areas, since the archaeology tends to be obscured by the tree canopies. By using airborne laser scanning data, also called airborne lidar data, and by only keeping the ground returns and not the returns from trees and buildings, the forest vegetation can be removed from the data, and a very detailed digital elevation model (DEM) of the ground surface can be constructed [1]. This makes it possible to detect

archaeology in a semi-automatic fashion, provided the archaeology manifests itself as features in the digital elevation model of the lidar ground returns, and that these features may be described using some appropriate kind of pattern. In the majority of Norway's 19 counties, there are intact grave mounds in forested areas. This means that a semi- automatic method for the detection of grave mounds in lidar data would be an important tool in a more systematic mapping of archaeology in Norway.

We have recently developed a method for the semi-automatic detection of hunting systems and iron extraction sites from airborne lidar data [7]. These archaeological features manifest themselves as pits in a digital elevation model (DEM) derived from the lidar ground returns. The method detects pits automatically in this DEM, followed by manual inspection by an archaeologist. This method is now in use as part of the standard procedure for archaeological mapping in Oppland County, Norway. This method can be modified to detect heaps instead of pits.

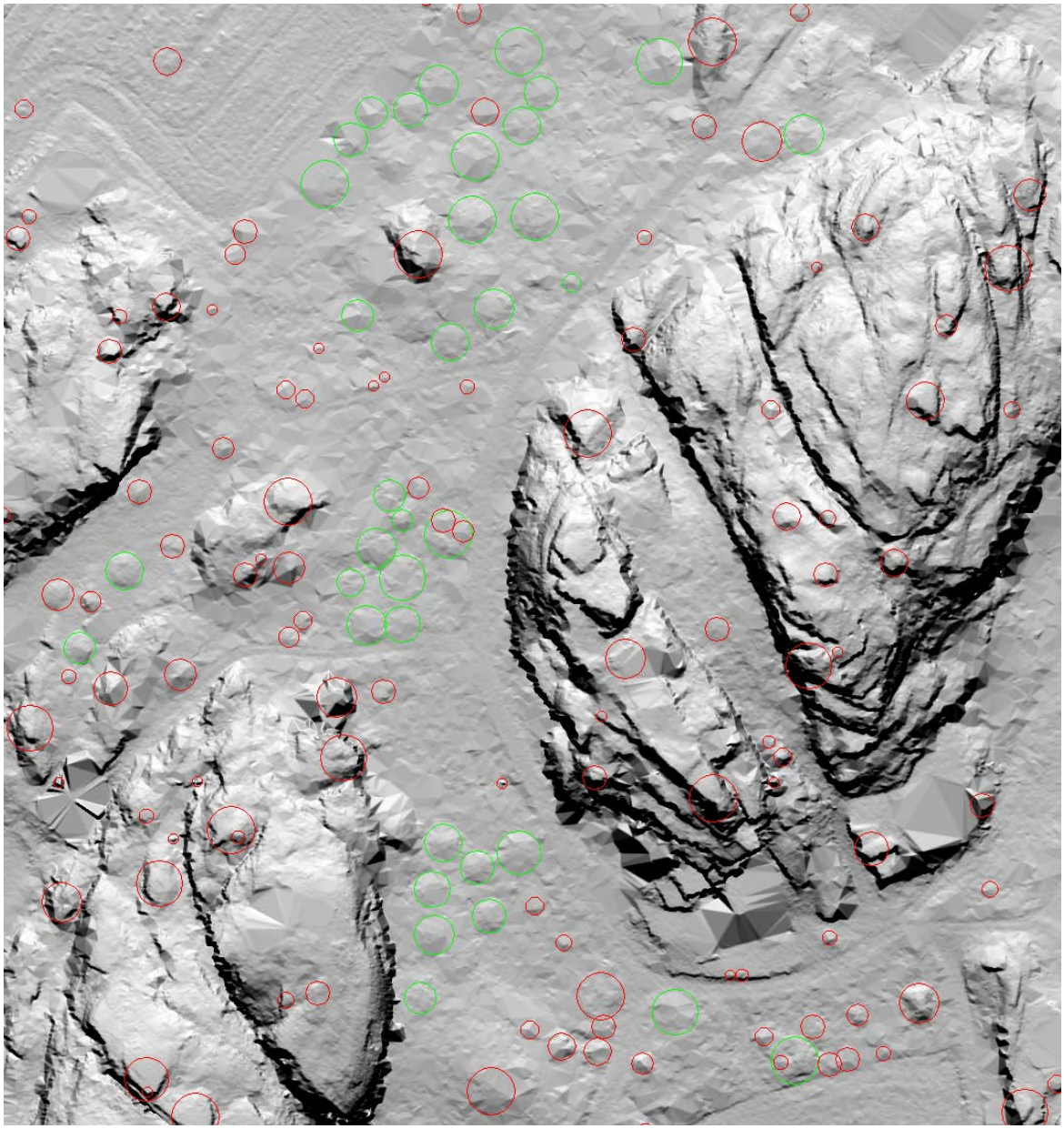

Figure 2. A 210 m × 225 m part of the Kaupang, Larvik training data set for heap detection. True (green) and false (red) grave mounds have been labelled manually.

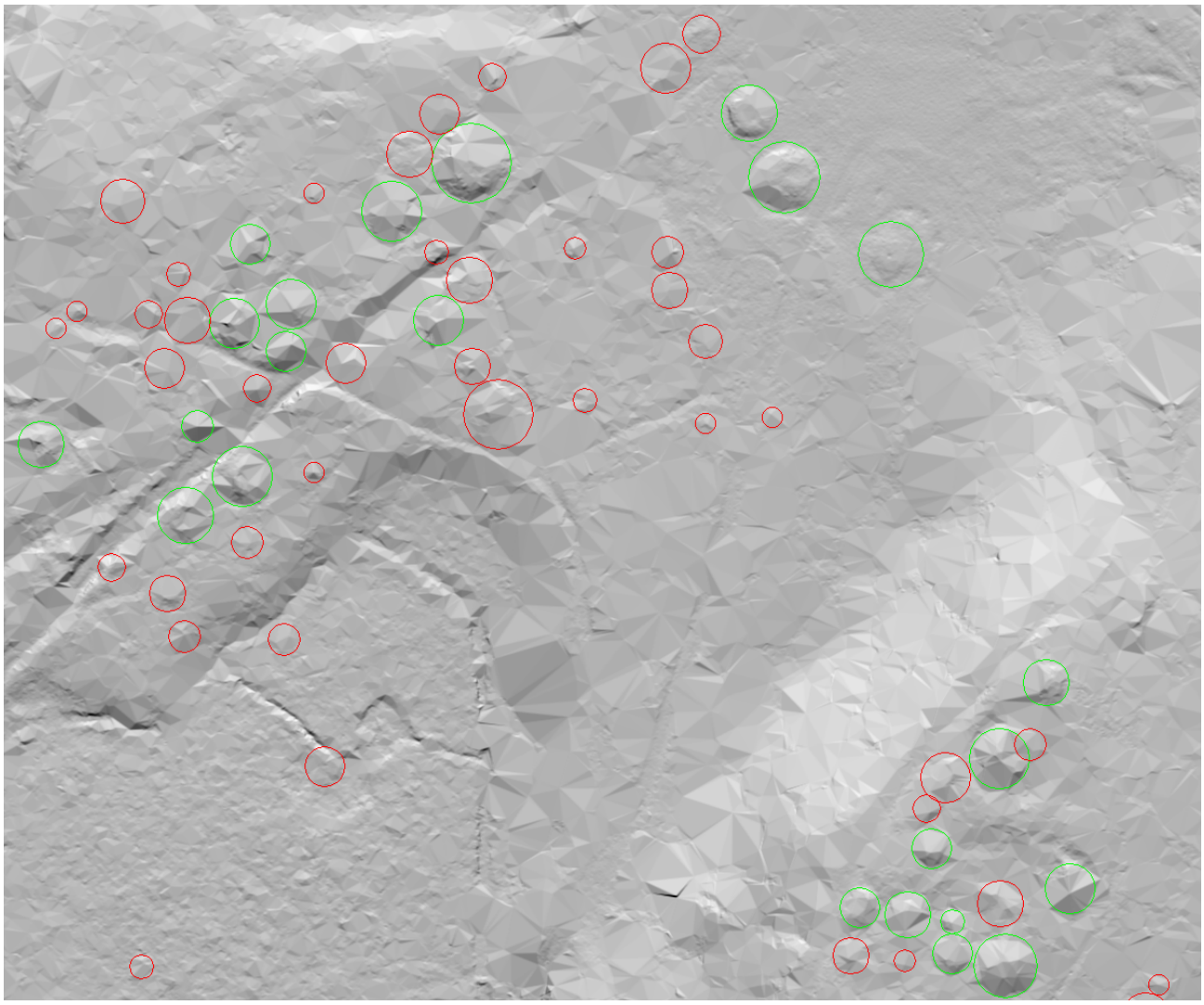

Figure 3. A 245 m × 200 m part of the Bøkeskogen, Larvik test data set for heap detection. True (green) and false (red) grave mounds have been labelled manually.

2. Data and Methods

2.1. Airborne lidar height measurements

Larvik municipality in Vestfold County is known to contain a large number of grave mounds in forested areas. A lidar data set covering about 150 km^2 of the southern part of Larvik municipality was acquired on 3-7 June 2010, with 22 emitted pulses per m^2 on average. From this data set, 12 small portions containing known grave mounds were extracted. Four of these are used as a training set: Kaupang (Figure 2), Store Sandnes, Tanum, and Ødelund. The remaining eight comprise a test set: Berg, Bommestad, Bøkeskogen (Figure 3), Hvatumskjeet, Kjerneberget, Lunde, Valby, and Valbysteinene.

2.2. Automatic detection of heap structures

Building on our previous work [6][7], we propose a processing chain for the automatic detection of heaps in lidar data:

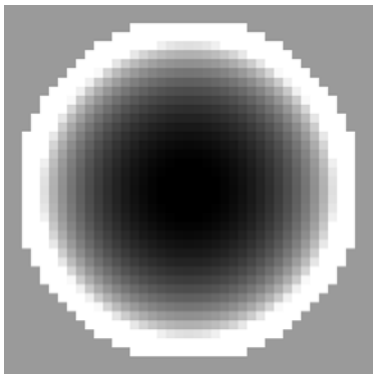

Figure 4. Heap template, shaped as a half-dome circumscribed by a flat ring. Black pixels are +1, white pixels are –1, and grey pixels in between. The medium grey pixels outside the white ring edge are exactly zero, thus not contributing to the convolution value. This particular heap template has 3.4 m radius

1. Obtain lidar data from a commercial provider, in the form of LAS files [3]. The point density should be at least 5 emitted pulses per m^2, and the discrete returns must be labelled as ground, building, vegetation, etc.
2. Convert the lidar ground returns to digital elevation models (height images) with 0.2 m resolution.
3. Convolve the height images with dome-shaped heap templates (Figure 4) of varying sizes. Threshold each convolution result to obtain candidate heap detections.
4. Merge detections that are overlapping, keeping the strongest detections.
5. For each candidate heap detection, compute various measures of the deviation from an ideal dome.
6. Using thresholds on some of the measures, remove the most obvious non-heaps.
7. Assign confidence levels, either using a statistical classifier, a decision tree classifier, or a combination.
8. The list of detected heaps is verified by an archaeologist, first by visual inspection of the lidar data, then by field work.

2.2.1. Computation of attributes

In step 5 above, the following 15 attributes are computed:

1. Correlation value, obtained from the convolution step.
2. Normalized correlation value, that is, the correlation value divided by the radius.
3. Average heap height, measured as the height difference between the highest point inside the heap and the average height on the ring edge outside the heap.
4. Minimum heap height, measured as the height difference between the highest point inside the pit and the highest point on the ring edge.
5. Normalized average heap height, that is, average heap height divided by the radius.
6. Normalized minimum heap height.
7. Standard deviation of height values on the ring edge.
8. Root mean square (RMS) deviation from a perfect hemisphere, i.e., a perfect U-shaped heap.
9. RMS deviation from a perfect V-shaped heap.
10. For each heap, a threshold is defined as the value that separates the pixels inside the heap into two groups, the 25% of the pixels that are brighter than the threshold, and the 75% that are darker. Use this threshold to extract a bright blob segment from a square image centred on the heap, with sides equal to six times the radius. This is called the 25%-segment. If this results in a compact, central segment inside the heap, connected to a

larger segment outside the pit, with only a few connecting pixels on a ring just outside the heap, then the central segment is separated from the outside segment. From the extracted segment, the following measures are computed:
   a. Offset: distance from heap centre to the segment's centre.
   b. Major axis length; for a definition, see e.g., [4].
   c. Elongation, defined as major axis divided by radius.
11. Similarly to above, extract the 50%-segment and compute offset, major axis and elongation from that segment as well.

2.2.2. Initial screening

Thresholds are set on some of the attributes to remove detections that are very unlikely to be archaeology, while at the same time keeping all true archaeological features. By sorting a training set of labelled detections on one attribute at a time, one can manually identify attributes that can be thresholded so that all detections labelled as 'true' or 'possible' archaeology be kept, keeping several 'unlikely' and 'false' detections as well, but at the same time removing many 'unlikely' and 'false' detections. These thresholds should not be set too tight, to allow for slightly more variation in the attribute values for the 'true' and 'possible' archaeological features than was observed in the training data.

2.2.3. Statistical classification versus decision tree

For step 7 in the detection method above, a manually designed decision tree could be used to assign confidence values 1-6, with 1 meaning 'very low' and 6 meaning 'very high' [7]. However, this requires that a number of fixed thresholds be set manually, based on training examples. An alternative is to use a statistical classifier, and use thresholds on the estimated probability that a heap is a grave mound. We have previously compared the two approaches for automatic pit detection in the context of semi-automatic detection of pitfall traps and charcoal burning pits [8]. The statistical approach was better than the manually constructed decision tree when the confidence was medium high or better. For medium or lower confidence, the manually constructed decision tree seemed to be better. Thus, it appears that a combined approach could be a good alternative. We will compare two approaches: (1) using statistical classifier, and (2) first using statistical classifier, then for confidence levels medium or lower, using a decision tree to reassign the confidence levels.

2.2.4. Automatic heap detection method: common steps

The first five steps in the heap detection method are common for both the manually designed decision tree and the statistical classifier approach. These five steps were applied on the Larvik data set. A number of parameters had to be selected in this process. A DEM grid size of 0.2 m was used to preserve the accuracy of the lidar height measurements, thus converting the ground hits to 25 interpolated height values per m^2. In the convolution step, heap templates corresponding to heap radii from 1.0 to 10.0 m were used, corresponding to the wide range of expected grave mound sizes. Each template has 0.2 m larger radius than the next smaller. As this is a first attempt, to avoid overlooking true grave mounds, the initial screening uses very relaxed thresholds on a subset of the attributes as follows:

1. Normalized correlation > 1.0
2. Average heap height > 0.2 m
3. Minimum heap height > 0.0 m
4. RMS u-shape < 0.2
5. RMS v-shape < 0.2
6. 25% segment elongation < 5

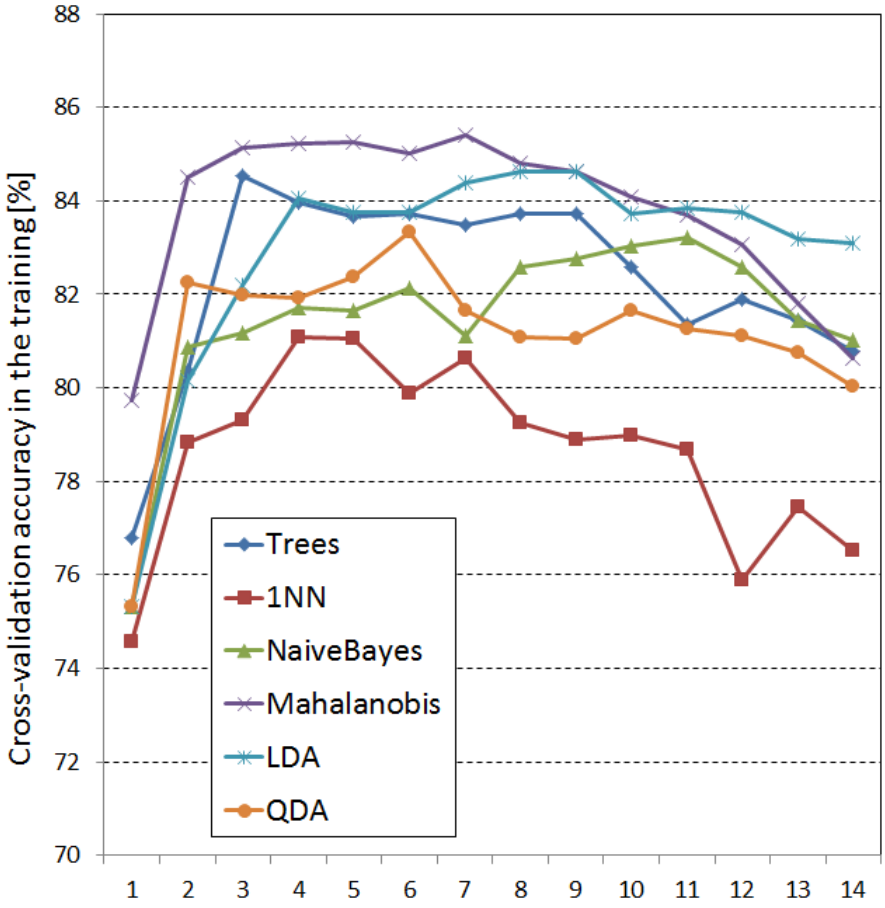

Figure 5. Performance of the six different classifiers on the Larvik training set, as a function of the number of attributes

The result of the initial screening was a training set with 785 heap detections, of which 96 were labelled 'true' and the remaining 'false'; and a test set of 905 heap detections, of which, 96 were labelled 'true' and the remaining labelled 'false'. The labelling was done by a non-archaeologist.

2.3. Automatic heap detection using statistical classifier

When using a statistical classifier, a number of parameters need to be estimated from the training data. The actual number of parameters varies between the different statistical classifiers. With a limited training set size, there is a trade-off. If we use a model with many parameters, less accurate estimates of these parameters may result than when using a model with fewer parameters. The following six different classifiers were evaluated [2]:

1. Decision tree (CART algorithm)
2. Nearest neighbour
3. Naïve Bayes (assuming independent attributes)
4. Mahalanobis distance
5. Linear discriminant analysis
6. Quadratic discriminant analysis

For each classifier, the best subset of the 15 attributes computed in Section 2.2.1 is determined using the sequential forward attribute selection algorithm [5]. The subset of attributes that maximizes the 10-fold cross-validation of average accuracy in the training set is retained. The best classifier turned out to be the Mahalanobis distance classifier (Figure 5), with the following seven attributes, in order of importance:

1. RMS U-shape
2. Correlation
3. Elongation of 25% segment
4. Offset of 25% segment
5. Standard deviation on edge
6. Major axis of 50% segment
7. Offset of 25% segment

We will now use the estimated posterior probability, that is, the probability that the detected heap is a grave mound, to assign a confidence level to each detection. With six confidence levels, we need to determine five thresholds. These can be found by defining and solving an optimization problem. As initial threshold values, we use the values corresponding to the 10th percentile, 25th, 50th, 75th and 90th percentile. Then we can count the number of pits and non-pits in each confidence level, multiply with penalty weights (Table 1) and accumulate to obtain a score for the particular choice of thresholds. The counts for the grave mound and non-grave mound classes are normalized according to the respective number of samples. By adjusting the threshold values iteratively using a sequential loop strategy, they can be optimized to minimize the final score. By doing this on the training set, the thresholds in Table 2 are obtained, which assign 'medium high' or better confidence to most of the true grave mounds, and 'medium' confidence or lower to most non-archaeological heaps (Table 3).

Table 1. Penalty weights used for optimizing confidence level thresholds.

score value	1	2	3	4	5	6
confidence	very low	low	medium	medium high	high	very high
pit	1024	256	64	16	4	1
non-pit	1	4	16	64	256	1024

Table 2. Optimized thresholds for confidence assignment for heap detection.

1	2	3	4	5
0.05072984	0.05121662	0.47666119	0.67167690	0.76737689

Table 3. The result of using the Mahalanobis distance classifier for confidence estimation on the Larvik training set.

score value	1	2	3	4	5	6	
confidence	very low	low	medium	medium high	high	very high	sum
grave mounds			16	47	21	12	96
not grave mounds	55	1	554	72	7		689
sum	55	1	570	119	28	12	785

2.3.1. Using a decision tree to reassign low confidence values

Since the statistical classifier does not assign meaningful confidence values when the confidence values are very low, low or medium, neither for pits, as observed in another study [8], nor for heaps (Table 3), a decision tree classifier may be used to reassign confidence levels medium or lower. For this purpose, a number of thresholds were used (Table 4).

Table 4 Thresholds for reassigning confidence levels

optimised on training set	confidence	
measurements	low	medium
RMS diff from U-shape	≤0.08	≤0.06
radius	≥1.4	≥1.8
correlation	≥1.2	≥2.0
25% segment elongation	≤4	≤1.6
standard deviation on ring edge	≤0.7	≤0.4
25% segment offset	≤25	≤15
normalized avgerage height	≤0.3	≤0.25
normalized avgerage height	≥0.05	≥0.06
normalized correlation	≤8	≤8
normalized correlation	≥1	≥1.5
average height	≥0.1	≥0.2

If a heap detected by the Mahalanobis distance classifier has medium or lower confidence, then it is first set to very low. If all the threshold tests for low in Table 4 are met, then the detection gets low confidence. Next, if all the thresholds for medium in Table 4 are met, then the confidence changes again from low to medium.

3. Results

3.1. Automatic heap detection using statistical classifier

By running the Mahalanobis distance classifier on the Larvik test set, a slightly higher number of true grave mounds obtained high or very high confidence (Table 5) compared with the training set (Table 3). At the same time, about twice as many false detections obtained medium high, high or very high confidence. None of the true detections and almost none of the false detections got ‘low’ or ‘very low’ confidence. So, in an operational setting, to successfully verify the 14 true grave mounds with ‘medium’ confidence, 647 false detections have to be checked as well. By accumulating the detection counts (Table 6), the trade-off between detecting as many grave mounds as possible while at the same time limiting the number of false detections is more evident. E.g., 82 out of 96 grave mounds were detected with medium high confidence or better (Table 6), this is 85% of the true grave mounds that were successfully segmented in the template matching step. At the same time, 158 of the detections with medium high or better confidence were false.

Table 5. Result of running the Mahalanobis distance classifier for confidence estimation on the Larvik test set

score value	1	2	3	4	5	6	
confidence	very low	low	medium	medium high	high	very high	sum
grave mound			14	39	25	18	96
not grave mound	4		647	144	13	1	809
sum	4	0	661	183	38	19	905

Table 6. Accumulated heap detection counts for the Mahalanobis distance classifier on the Larvik test set.

score value	≥1	≥2	≥3	≥4	≥5	≥6	
confidence	very low or better	low or better	medium or better	medium high or better	high or better	very high	sum
grave mound	**96**	**96**	**96**	**82**	**43**	**18**	**96**
not grave mound	809	805	805	158	14	1	809
sum	905	901	901	240	57	19	905
grave mounds detected	100 %	100 %	100 %	85 %	45 %	19 %	
grave mounds missed	0 %	0 %	0 %	15 %	55 %	81 %	

Table 7. Result of running the combined classifier for confidence estimation on the Larvik test set.

confidence	very low	low	medium	medium high	high	very high	sum
grave mound	1	5	8	39	25	18	96
not grave mound	145	351	155	144	13	1	809
sum	146	356	163	183	38	19	905

Table 8. Accumulated heap detection counts for the combined classifier on the Larvik test set.

confidence	very low or better	low or better	medium or better	medium high or better	high or better	very high	sum
grave mound	**96**	**95**	**90**	**82**	**43**	**18**	**96**
not grave mound	809	664	313	158	14	1	809
sum	905	759	403	240	57	19	905
grave mounds detected	100 %	99 %	94 %	85 %	45 %	19 %	
grave mounds missed	0 %	1 %	6 %	15 %	55 %	81 %	

In order to obtain more true detections, e.g., by considering all detections with medium or better confidence, 805 false detections are obtained as well. By using the decision tree to reassign confidence levels for detections with medium or lower confidence, the same numbers of true and false detections with medium high confidence as before are obtained (Table 7). However, it is now possible to detect eight additional grave mounds, with medium confidence, with the additional cost of obtaining 155 false detections, labelled with medium confidence. This means that 94% of the true grave mounds are detected with medium confidence or better, with 313 false detections in addition (Table 8).

When running the automatic method on the Kaupang part of the training data, and overlaying with grave monuments from the official "Askeladden" Norwegian cultural heritage database, it becomes evident that a number of true grave mounds are not detected by the method (Figure 6). The percentages in Table 5-Table 8 are estimated without taking these missing detections into account.

4. Discussion and conclusions

The heap detection results on the Larvik test set indicate that the combined confidence assignment method is capable of assigning medium confidence or better to 94% of the grave mounds that it is applied on, while at the same time, 3-4 times as many false detections as true detections are assigned medium confidence or better. From previous experience [7], this is an acceptable trade-off. However, a substantial number of grave mounds have not been detected at all

(Figure 6). This could either be due to missed detections in the template matching step, or due to removal of true detections in the subsequent thresholding step, which aims at removing obvious non-heaps. We need to investigate the reason these were missed, in order to improve the method.

Further, we need to quantify how the method performs in areas with few grave mounds. We suspect that the number of false detections remains relatively stable for geographical areas of similar size, which means that the number of false detections could be overwhelming. If this is indeed the case, we need to consider ways of improving the method.

However, improving the method is of limited value if the problems are due to the quality of the lidar data. Previous experience with pit detection in lidar data clearly demonstrates the negative effect of reduced ground point density. Low ground point density may be due to wrong acquisition time, dense or low vegetation, too few emitted pulses per square meter, or a combination of these. Since there is a large proportion of deciduous trees in the forests in Larvik municipality, the archaeologists recommended that the acquisition of lidar data be done in late April 2010, which would have been an ideal time, with no snow on the ground and no leaves on the trees. For some reason, the lidar acquisition was postponed until the beginning of June 2010, with full-size leaves on the trees. We recommend that future lidar data sets be acquired in the early spring with no leaves on the deciduous trees, to allow more true grave mounds to be detected by the method.

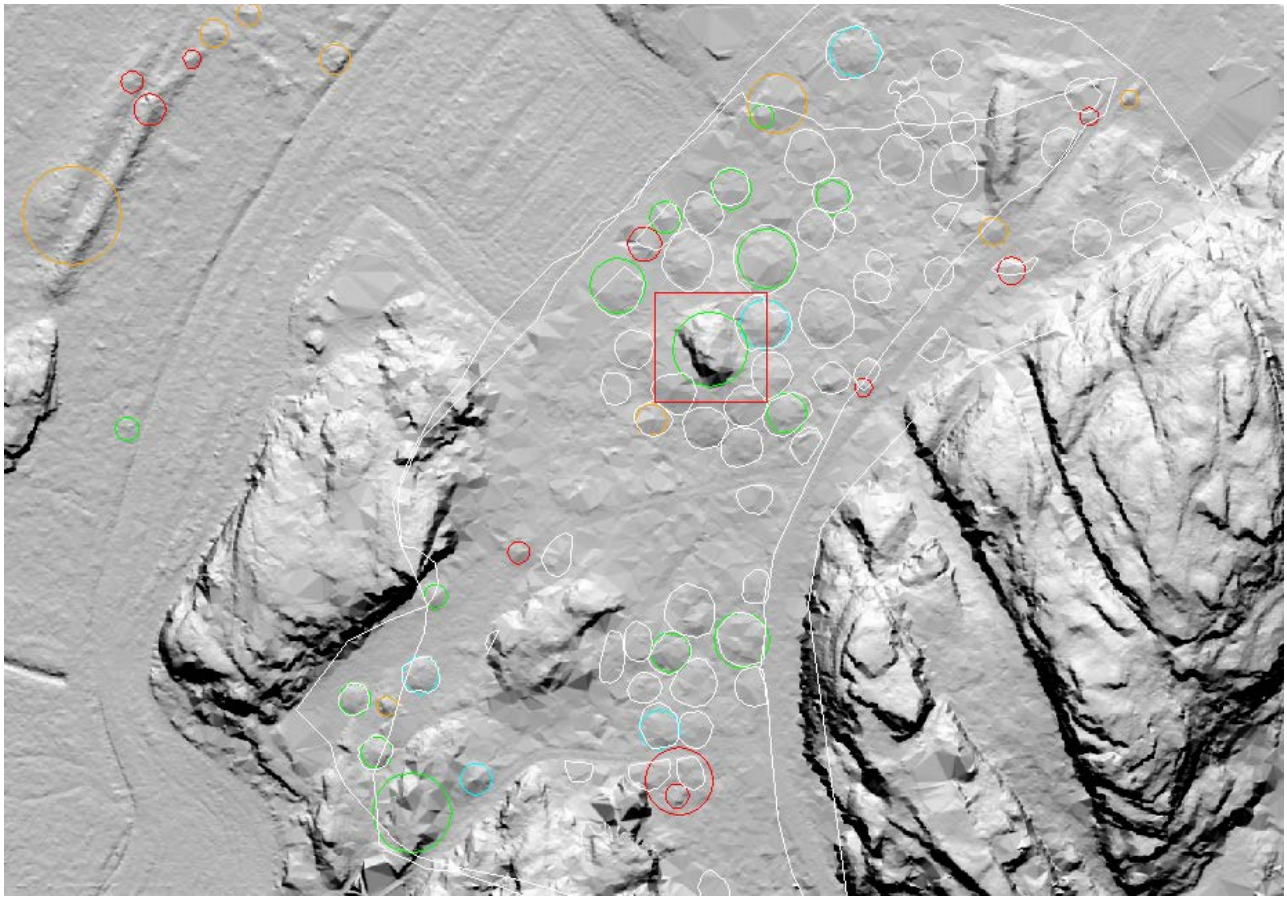

Figure 6. Detection results on a part of the Kaupang training data. The white polygons are previously mapped grave monuments, containing individually mapped grave mounds and grave field boundaries. Coloured circles are automatic heap detections, using the combined method for confidence assignment: blue=very high, cyan=high, green=medium high, yellow=medium, orange=low, red=very low. The red square indicates a natural terrain feature which has been detected as a grave mound with medium high confidence (green ring).

Another potential problem related to the lidar ground point density is that some grave mounds are small, either measured as height difference relative to the surrounding landscape, or measured as radius from centre to edge. Clearly, if the grave mound does not manifest itself in the data, then it cannot be detected. There could also be a problem if the heap template does not resemble the true shape of grave mounds. One possible way of studying this is to estimate from the data the shape of an average grave mound, and how it varies; then to generate simulated grave mound templates.

Finally, we could consider extracting more attributes from the detected heap candidates, in the hope that some of these may improve the ability to discriminate between grave mounds and non-archaeological heaps. Additional attributes could include, e.g.:

1. Average lidar ground point density within the area covered by the heap template
2. Average lidar intensity
3. Average height gradient
4. Average gradient squared
5. Average gradient entropy.

The motivation for the various gradient measurements is that some heap detections are natural terrain features, which, by inspection of the lidar data, contain steeper slopes than found on grave mounds. Two of the suggested gradient measures are weighted averages, which place more emphasis on the high gradient values than a non-weighted average. The intensity might give some hints as to the hardness of the ground surface, as well as whether part of the emitted pulse did not reach the ground. The point density could indicate how well the shape of a possible heap is preserved.

As a conclusion, the present study demonstrates that automatic heap detection could be a useful tool for the semi-automatic detection of grave mounds in Norway from airborne laser scanning data, provided the number of ground points per square meter is not too low. The method needs improvement to be used in an operational setting with non-optimal data.

Acknowledgements

We thank Vestfold County Administration for providing lidar data, and the Norwegian Directorate for Cultural Heritage for funding the project.

References

[1] Devereux, B. J., Amable, G. S., Crow, P., Cliff, A. D., 2005. The potential of airborne lidar for detection of archaeological features under woodland canopies. Antiquity 79, pp. 648-660.

[2] Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning. Data mining, inference and prediction*. Second edition. Springer, New York.

[3] *LAS specification*, version 1.3 – R11, October 24, 2010. The American Society for Photogrammetry & Remote Sensing, 18 pp. [online 2012-02-07] URL: http://www.asprs.org/a/society/committees/standards/LAS_1_3_r11.pdf

[4] Prokop, R. J., Reeves, A. P. 1992. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing* 54(5), pp. 438–460.

[5] Pudil, P., Novovičova, J., Kittler, J., 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15, pp. 1119-1125.

[6] Trier, Ø. D., Larsen, S. Ø., Solberg, R., 2009. Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection* 16(1), pp. 1-15. DOI: 10.1002/arp.339.

[7] Trier, Ø. D., Pilø, L. H., 2012. Automatic detection of pit structures in airborne laser scanning data. *Archaeological Prospection*, to appear.

[8] Trier, Ø. D., Zortea, M., 2012. Semi-automatic detection of cultural heritage in lidar data. In *4th International Conference on Geographic Object-Based Image Analysis (GEOBIA)*, 7-9 May 2012, Rio de Janeiro, Brazil.