Semi-automatic detection of burial mounds in forested areas

Øivind Due Trier, Maciel Zortea and Siri Øyen Larsen

EARSeL 32nd Annual Symposium
Mykonos, Greece, 21-24 May 2012
Grave mounds
Grave mounds
Semi-automatic detection of archaeology in Norway

► Traditional mapping of cultural heritage in Norway is inaccurate and incomplete

► Large amounts of very high resolution remote sensing data are being collected
  ▪ Lidar
  ▪ Aerial photographs
  ▪ Satellite imagery, e.g., Worldview-2, Quickbird

► Could semi-automatic methods be used on these datasets to map archaeology more accurate and complete?

► Project with Riksantikvaren (Norwegian Directorate for Cultural Heritage) since 2002
Success: mapping pitfall traps in lidar data: Olstappen, Nord-Fron municipality, Norway
Success: Mapping of archaeological pits in lidar data: Øystre Slidre

From field survey:

- Single coal pit
- Iron production
- Not cultural heritage

Automatic detection, with confidence:

- 5 – high
- 4 – medium high
- 3 – medium
- 2 – low

Not inspected in the field:

- Not cultural heritage

GPS track

Visual detection
Benefits of semi-automatic detection

► Fast processing of huge amounts of data
► Accurate measurements of position and size
► Provides an initial mapping for subsequent field survey
► Hunting systems and iron extraction sites in Oppland County:
  ▪ The combined use of automatic detection and visual inspection of lidar data, prior to field survey, makes the field survey at least 10 times faster than the traditional method.
Heap detection challenges

- Grave mounds may be less distinct in the lidar data
- Discriminate between archaeological heaps, modern man-made heaps, and natural terrain features

- Relevant object properties

- Classifier

- Lidar point density on the ground
  - Acquisition time (leaves on/off)
  - Emitted pulses per m²
  - Vegetation density, low vegetation
Pattern recognition method

Assumes that objects to be detected may be described using some sort of pattern

1. Identify heap candidates (template matching)
2. Measure properties of heap candidates
3. Classify heap candidates
   1. Heap vs non-heap
   2. Confidence of heap being archaeology
4. Visual inspection
5. Field survey
Properties of objects

In use:

- correlation
- radius
- correlation / radius
- minimum height
- average height
- minimum height / radius
- average height / radius
- standard deviation on ring edge
- root mean square (RMS) difference from U-shape
- RMS diff from V-shape
- 50% segment offset
- 50% segment major axis
- 50% segment elongation
- 25% segment offset
- 25%-segment major axis
- 25%-segment elongation

Will investigate:

- Average point density
- Average intensity
- Average gradient
- Squared gradient
- Gradient entropy
- Combinations
What is the 25%-segment? Example with pits

- Threshold = limit for the 25% darkest pixels inside pit (of radius r)
- Threshold 6r x 6r subimage centered on pit

<table>
<thead>
<tr>
<th>Pitfall trap</th>
<th>Pitfall trap</th>
<th>Pitfall trap</th>
<th>Pitfall trap</th>
<th>Road edge</th>
<th>Valley</th>
<th>Valleys meet</th>
<th>Road edge</th>
<th>Foothill</th>
<th>Rock in slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major axis</td>
<td>11.04</td>
<td>12.00</td>
<td>12.68</td>
<td>8.52</td>
<td>32.91</td>
<td>40.87</td>
<td>61.51</td>
<td>28.19</td>
<td>88.95</td>
</tr>
<tr>
<td>Radius</td>
<td>12.00</td>
<td>12.00</td>
<td>11.00</td>
<td>6.00</td>
<td>16.00</td>
<td>12.00</td>
<td>17.00</td>
<td>6.00</td>
<td>17.00</td>
</tr>
<tr>
<td>Elongation</td>
<td>0.92</td>
<td>1.00</td>
<td>1.15</td>
<td>1.42</td>
<td>2.06</td>
<td>3.41</td>
<td>3.62</td>
<td>4.70</td>
<td>5.23</td>
</tr>
</tbody>
</table>
Classifier

► Initial screening: Remove obvious non-heaps based on fixed thresholds

► Confidence assignment:
  - Very low
  - Low
  - Medium
  - Medium high
  - High
  - Very high

► Manually set thresholds vs. statistical classifier
Confidence values

- Tight or loose thresholds?
  - Number of missed heaps vs. number of false detections

- A confidence value on each detection reflects this

- Example thresholds and rules:

<table>
<thead>
<tr>
<th>example thresholds</th>
<th>very low</th>
<th>low</th>
<th>medium</th>
<th>med. high</th>
<th>high</th>
<th>very high*</th>
</tr>
</thead>
<tbody>
<tr>
<td>measurements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>normalized correlation</td>
<td>≥1</td>
<td>≥2</td>
<td>≥2.5</td>
<td>≥3.0</td>
<td>≥3.5</td>
<td></td>
</tr>
<tr>
<td>minimum height</td>
<td>≥0.05</td>
<td>≥0.1</td>
<td>≥0.15</td>
<td>≥0.25</td>
<td>≥0.4</td>
<td>≥1</td>
</tr>
<tr>
<td>average height</td>
<td>≥0.25</td>
<td>≥0.4</td>
<td>≥0.45</td>
<td>≥0.5</td>
<td>≥0.55</td>
<td></td>
</tr>
<tr>
<td>RMS U-shape</td>
<td>≤0.2</td>
<td>≤0.1</td>
<td>≤0.09</td>
<td>≤0.08</td>
<td>≤0.07</td>
<td>≤0.02</td>
</tr>
<tr>
<td>RMS V-shape</td>
<td>≤0.2</td>
<td>≤0.1</td>
<td>≤0.08</td>
<td>≤0.07</td>
<td>≤0.05</td>
<td>≤0.015</td>
</tr>
<tr>
<td>25% segment offset</td>
<td>≤20</td>
<td>≤10</td>
<td>≤8</td>
<td>≤6</td>
<td>≤4</td>
<td></td>
</tr>
<tr>
<td>25% segment elongation</td>
<td>≤4</td>
<td>≤2</td>
<td>≤1.75</td>
<td>≤1.5</td>
<td>≤1.25</td>
<td></td>
</tr>
<tr>
<td>assigned tag</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>
Six different classifiers are evaluated:
- Decision tree (CART algorithm)
- Nearest neighbour
- Naïve Bayes (assuming independent attributes)
- Mahalanobis distance
- Linear discriminant analysis
- Quadratic discriminant analysis

For each classifier, find the best subset of attributes
Evaluation of statistical classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>Accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>13</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>14</td>
<td>6</td>
<td>84.5</td>
</tr>
<tr>
<td>1NN</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>14</td>
<td>10</td>
<td>13</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>1</td>
<td>9</td>
<td>81.1</td>
</tr>
<tr>
<td>NaiveBayes</td>
<td>13</td>
<td>14</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>15</td>
<td>10</td>
<td>7</td>
<td>12</td>
<td>5</td>
<td>2</td>
<td>11</td>
<td>4</td>
<td>8</td>
<td>83.2</td>
</tr>
<tr>
<td>Mahalanobis</td>
<td>7</td>
<td>1</td>
<td>14</td>
<td>12</td>
<td>6</td>
<td>10</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>85.4</td>
</tr>
<tr>
<td>LDA</td>
<td>7</td>
<td>13</td>
<td>6</td>
<td>14</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>12</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>84.6</td>
</tr>
<tr>
<td>QDA</td>
<td>13</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>83.3</td>
</tr>
</tbody>
</table>

Attributes of heaps:

1 = correlation
2 = minimum height
3 = average height
4 = normalized minimum height
5 = normalized average height
6 = standard deviation on ring edge
7 = root mean square (RMS) difference from U-shape
8 = RMS diff from V-shape
9 = 50% segment offset
10 = 50% segment major axis
11 = 50% segment elongation
12 = 25% segment offset
13 = 25%-segment major axis
14 = 25%-segment elongation
Evaluation of statistical classifiers

Cross-validation accuracy in the training [%]

- Trees
- 1NN
- NaiveBayes
- Mahalanobis
- LDA
- QDA

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Confidence estimation with statistical classifier

- Statistical classifier: the probability that an object is a grave mound
- A confidence level may be obtained by thresholding this probability
- Initial thresholds

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>

- Penalty weights for misclassifications

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>confidence</td>
<td>very low</td>
<td>low</td>
<td>medium</td>
<td>medium high</td>
<td>high</td>
<td>very high</td>
</tr>
<tr>
<td>pit</td>
<td>1024</td>
<td>256</td>
<td>64</td>
<td>16</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>non-pit</td>
<td>1</td>
<td>4</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
</tr>
</tbody>
</table>

- Optimize thresholds to minimize accumulated penalty on training data
- Final thresholds

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05072984</td>
<td>0.05121662</td>
<td>0.47666119</td>
<td>0.67167690</td>
<td>0.76737689</td>
<td></td>
</tr>
</tbody>
</table>
Performance on heaps (Larvik)

<table>
<thead>
<tr>
<th>confidence</th>
<th>very low</th>
<th>low</th>
<th>medium</th>
<th>medium high</th>
<th>high</th>
<th>very high</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>grave mound</td>
<td>4</td>
<td>647</td>
<td>14</td>
<td>39</td>
<td>25</td>
<td>18</td>
<td>96</td>
</tr>
<tr>
<td>not grave mound</td>
<td>0</td>
<td>0</td>
<td>661</td>
<td>183</td>
<td>38</td>
<td>19</td>
<td>809</td>
</tr>
<tr>
<td>sum</td>
<td>4</td>
<td>0</td>
<td>661</td>
<td>183</td>
<td>38</td>
<td>19</td>
<td>905</td>
</tr>
</tbody>
</table>

- Statistical method not able to discriminate between very low, low and medium confidence
- Could combine manual thresholds for very low – medium confidence with statistical classifier for medium high – very high confidence
- Alternatives:
  - manually adjust thresholds on posterior probability from statistical classifier
  - Lower punishment weights for true grave mounds which receive medium or low confidence
Combined classifier

- Use statistical classifier first
- If confidence <= medium, then reassign confidence using manually set thresholds

<table>
<thead>
<tr>
<th>confidence</th>
<th>very low</th>
<th>low</th>
<th>medium</th>
<th>medium high</th>
<th>high</th>
<th>very high</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>grave mound</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>39</td>
<td>25</td>
<td>18</td>
<td>96</td>
</tr>
<tr>
<td>not grave mound</td>
<td>145</td>
<td>351</td>
<td>155</td>
<td>144</td>
<td>13</td>
<td>1</td>
<td>809</td>
</tr>
<tr>
<td>sum</td>
<td>146</td>
<td>356</td>
<td>163</td>
<td>183</td>
<td>38</td>
<td>19</td>
<td>905</td>
</tr>
</tbody>
</table>
**Manual thresholds for heaps**

<table>
<thead>
<tr>
<th>Optimised on training set</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurements</td>
<td>Low</td>
</tr>
<tr>
<td>RMS diff from U-shape</td>
<td>≤0.08</td>
</tr>
<tr>
<td>Radius</td>
<td>≥1.4</td>
</tr>
<tr>
<td>Correlation</td>
<td>≥1.2</td>
</tr>
<tr>
<td>25% segment elongation</td>
<td>≤4</td>
</tr>
<tr>
<td>Standard deviation on ring edge</td>
<td>≤0.7</td>
</tr>
<tr>
<td>25% segment offset</td>
<td>≤25</td>
</tr>
<tr>
<td>Normalized average height</td>
<td>≤0.3</td>
</tr>
<tr>
<td>Normalized average height</td>
<td>≥0.05</td>
</tr>
<tr>
<td>Normalized correlation</td>
<td>≤8</td>
</tr>
<tr>
<td>Normalized correlation</td>
<td>≥1</td>
</tr>
<tr>
<td>Average height</td>
<td>≥0.1</td>
</tr>
</tbody>
</table>
Detection results: Kaupang (Larvik)
Detection results: Ødelund (Larvik)
Gradient

- Can high gradient indicate non-archaeological heap?
- Currently investigating several gradient measures:
  - Sobel gradient $g$
  - Gradient squared $g^2$
  - Gradient entropy $\sqrt{g} \cdot g$
What can go wrong?

► Too few lidar ground points per m²
  ▪ Wrong acquisition time
  ▪ Dense vegetation or low vegetation
  ▪ Too few emitted pulses per m²

► Heap is small (height and/or radius)

► Template matching fails to detect all grave mounds

► Heap measurements are not useful in separating grave mounds from natural heaps and modern man-made heaps

► Classifier
  ▪ Initial screening removes some true grave mounds
  ▪ Confidence assignment
Conclusions

► Automatic detection of pits in lidar data has become an important tool as part of archaeological survey of cultural heritage in Oppland County, Norway

► Similar methods are needed for the automatic detection of heaps in lidar data for the mapping of grave mounds

► The current method is promising but needs further improvement
Alternative conclusions

- Bad data => bad results
- Good data => good results
Also in the project: mapping of levelled grave mounds

- Cropmarks
- Georadar
- Digging
Acknowledgements

► This project is sponsored by the Norwegian Directorate for Cultural Heritage (in Norwegian: Riksantikvaren)

► Fieldwork in the project has been conducted by Oppland County, Vestfold County, The Museum of Cultural History at the University of Oslo, and The Norwegian Institute for Cultural Heritage Research (NIKU)
Thank you for your attention!