The data explosion: tackling the taboo of automatic feature recognition in airborne survey data

Rebecca Bennett[1], Dave Cowley[2] & Véronique De Laet[3,4]

Introduction

The increasing availability of multi-dimensional remote-sensing data covering large geographical areas is generating a new wave of landscape-scale research that promises to be as revolutionary as the application of aerial photographic survey during the twentieth century. Data are becoming available to historic environment professionals at higher resolution, greater frequency of acquisition and lower cost than ever before. To take advantage of this explosion of data, however, a paradigm change is needed in the methods used routinely to evaluate aerial imagery and interpret archaeological evidence. Central to this is a fuller engagement with computer-aided methods of feature detection as a viable way to analyse airborne and satellite data. Embracing the new generation of vast datasets requires reassessment of established workflows and greater understanding of the different types of information that may be generated using computer-aided methods.

Automated and semi-automated image analysis is routine in fields such as environmental remote sensing, where it underpins analysis of extensive datasets (see Lasaponara & Masini 2012 for a review of these techniques). Aspects of these developments have made their way into archaeological applications, but they remain rarely used and are often viewed with suspicion. It is important therefore to outline the *status quo* and identify the key issues in debates that have mainly played out at conferences and steering groups but are seldom committed to print. There are difficulties in the creation of geographically extensive, systematic datasets for heritage management and research from high-volume data derived from airborne laser scanning (ALS), satellite imagery and airborne digital spectral sensors. Yet fuller exploitation of computer assisted techniques can exponentially increase the rapidity with which initial historic environment datasets can be created, moving beyond the 'human-timescales' within which most archaeological survey is undertaken (i.e. 'manual' analysis). This requires recognition of the fundamentally different but complimentary types of information that the two approaches produce, and assessment of their value and contribution as part of broader research objectives. Throughout we use the term 'computer vision' to define all methods by which imagery can be processed, analysed and

1 *Department of Archaeology, University of Winchester, King Alfred Campus, Sparkford Road, Winchester SO22 4NR, UK*

2 *Royal Commission on the Ancient and Historical Monuments of Scotland, John Sinclair House, 16 Bernard Terrace, Edinburgh EH8 9NX, UK*

3 *Department of Earth and Environmental Sciences, KU Leuven, Celestijnenlaan 200E, B-3001 Leuven-Heverlee, Belgium*

4 *3GeoID, Researchpark Haasrode, Interleuvenlaan 62, 3001 Leuven, Belgium*

ANTIQUITY 88 (2014): 896–905 http://antiquity.ac.uk/ant/088/ant0880896.htm

understood using computer-generated algorithms (including classifications), with reference to the electronic replication of the abilities of human ocular perception (Sonka *et al*. 2008).

Discussion of computer vision and pattern detection techniques in archaeology has often been from polarised positions that are as much about belief in the value of a particular type of output as a rational assessment of the contribution of new techniques. There is little doubt that the rhetoric has softened in recent years, thanks in part to the increasingly mainstream use of digital images rendered from ALS data, but also to a number of considered and rigorous applications of automated techniques (e.g De Laet *et al*. 2007; Trier & Pilø 2012; Verhagen & Drăguţ 2012; Lambers & Zingman 2013). There is, however, still vocal opposition within historic environment circles to any type of computer-generated detection process, stemming from a view that computer vision techniques are no substitute for manual interpretation (Parcak 2009; Hanson 2010). While recognising the legitimacy of some concerns, we argue that most objections overlook the potential role of such techniques as complementary to, rather than a direct substitute for, traditional methods.

A twentieth-century approach to a twenty-first-century problem

A sound and systematic knowledge base is fundamental to effective archaeological research and heritage management (Horne 2009; Cowley & Huld Sigurðardóttir 2011). Globally, however, the historic environment is poorly recorded both in terms of spatial coverage and quality of information. In parts of Europe with long traditions of inventory going back into the nineteenth century, analysis of the contents and structure of the records quickly illustrates the unsystematic nature of data collection and record compilation, alongside many inherent biases (Cowley 2011; Risbøl *et al*. 2013). While techniques such as aerial photographic analysis have revolutionised archaeological knowledge in some areas (e.g. Gojda 2011), these depend on manual and time-consuming assessment processes to create baseline records.

The ability of historic environment professionals to engage with large datasets is impeded by an overwhelming adherence to an entirely manual prospection and interpretation approach, as traditionally applied to aerial photographic interpretation (Wilson 2000; Cowley *et al*. 2013). While this can produce detailed, research-engaged interpretations of the data, it is also time consuming and requires heavy commitment of resources. Thus the ground-breaking English Heritage National Mapping Programme (Horne 2009), which draws principally on aerial photographs, achieves an average coverage rate of 1km^2 per person/day. With a variable number of staff (averaging 15–20 for long periods), the programme has examined about 52 000km^2 of English countryside since its inception in the 1990s, locating over 100 000 archaeological sites. This project has created records that underpin effective heritage management and support better understanding of the past, but its exclusively manual approach requires significant resources. With accelerating rates of landscape change and the limited resourcing of cultural heritage globally, traditional methods of remote landscape survey are a financially unsustainable and painfully inefficient way to create the spatially extensive, systematic datasets required for effective resource management.

Additional problems are encountered when traditional methods are applied to new datasets. Multiple methods of visualisation are often required to render archaeological features from topographic data (Bennett *et al.* 2012), while modern air- and spaceborne spectral systems have such high spectral dimensionality that a single survey results in hundreds of images to be assessed. Extracting relevant content from these images exceeds what is practically possible for a traditional manual approach and moreover this methodology fails to engage with the depth of content that such data has to offer. For example, the additional spectral response information could be critical to understanding how and why features of archaeological interest manifest in the imagery (Beck 2011; Bennett *et al.* 2011; Verhoeven 2012). There is a fundamental tension generated by the inability of traditional manual approaches to enable archaeological interpretation of extensive and complex remote sensing data. The pressure to break from tradition and explore other methods has undoubtedly contributed to the polarised debate on the role of automatic feature identification and computer vision in archaeological landscape analysis. However, there is potential for a middle ground that recognises the role of the expert in creating reliable interpretations, but looks to the development of analytical frameworks that are appropriate to the emergent datasets.

The role of the expert

A distinct advantage of manual interpretation is the ability to identify and categorise the wide range of direct and proxy features that are of potential archaeological significance, and skilled archaeological interpreters are crucial to the quality of feature interpretation (Palmer 2011; Halliday 2013). Thus, many believe that any form of semi-automated assessment is of little worth in comparison to knowledge-based expert interpretation. Parcak (2009: 110–11) promotes this view, arguing that computers simply do not have the ability of the human eye and mind to detect patterns that might denote traces of past human activity and separate them from geological and recent disturbances. The sophisticated use of image processing and artefact recognition in a wide range of applications from number plate and face recognition to medical imaging and military reconnaissance would undermine this view, but it finds some support in the history of automated detection techniques in archaeology. New remote sensing technologies for archaeological prospection have had to rely on data and image processing expertise from other fields. This has undeniably led to poor communication of purpose and over-enthusiastic claims of success, particularly as the type of features sought by archaeologists are categorically different from those the environmental and geological remote-sensing communities are trained to observe. Detecting one feature or one class of feature in one type of environment is not the same as providing both identification and interpretation of features across a landscape. This has fed criticism of automated detection focusing on high rates of 'false positives' (Hanson 2010), and generated fears that expertise in archaeological interpretation is undervalued.

The importance of interpretation based on the cognitive abilities, expertise and experience of the interpreter should not be overlooked. The complex mix of observation, feature identification, classification, interpretation and narrative are virtually impossible to disambiguate, and it is important to recognise that they are rarely examined or held to account (Halliday 2013). Deconstructing these often subconscious processes can help us to better understand the cognitive processes and reasoning underlying traditional approaches.

Archaeological interpretations are heavily conditioned by varying observational abilities, experience and knowledge. Our innate ability to identify patterns, which is a strength in analysing visual data, can sometimes mislead. There is the widely recognised tendency to 'complete the circle' where features are 'fitted' to the template of a familiar shape. Expectation and our existing knowledge-base will bias what we see, or do not see, and how we interpret it (Figure 1): "give half a dozen people the same information . . . you get half a dozen different interpretations, all equally valid and probably not even mutually contradictory in any significant way, but each of them will bear the style of the individual interpreter" (Hill 2009: 3). Multiplicity of interpretations is inevitable and yet often undocumented, leading to a tendency to view an expert transcription as the paramount and only interpretation of the data.

However, experienced interpreters would admit that the information they derive is heavily biased towards their expectations and knowledge, with the risk of missing or dismissing features (Halliday 2013). If one subscribes even partially to this conditional and contingent view of interpretation, then it seems unwise to argue for the primacy or exclusivity of human or manual processes with no room for computer vision techniques.

A computer algorithm designed to extract features with specific attributes from data is not as flexible as a human observer, and is therefore not able to filter and rationalise a mass of visual information. While the algorithm is not flexible, it removes a major source of bias in detection and ensures that all features within the set criteria will be detected. The value of information extracted will depend on the coherence of the enquiry, but it will be systematic and can be applied across large datasets, and, moreover, with good documentation it can be replicated. Such an approach does not, however, provide archaeological interpretations. This is an important point because some critics of automated and computer vision approaches have conflated feature detection and archaeological interpretation, largely because in a traditional manual workflow they are invariably intertwined. With the systematic application of explicit algorithms to identify features, a major source of bias in observation and detection is offset, with the potential to flag up features that might have been dismissed by the observer. Moreover, a growing number of researchers are developing analyses in an iterative process which takes account of field observation and evaluation to assess and refine the algorithms (Trier & Pilø 2012; Lambers & Zingman 2013).

Recognising the value of different approaches

Critique of manual and semi-automated methods for feature detection is hampered by the fact that the two methods result in different products and therefore cannot easily be compared. This can be demonstrated by the results of two aerial prospection projects with different methodologies (Table 1). The coverage rate of the Baden-Württemberg project is impressive compared to the NMP approach, but the results are considered 'potential archaeological features' that require verification. The Baden-Württemberg product is seen as a baseline, a segmented image highlighting areas of potential archaeological features based on predetermined rules for subsequent verification. The NMP approach adds to feature detection a level of archaeological interpretation that allows for assessment of significance and meaning; in doing so it also rejects without record those features deemed insufficiently

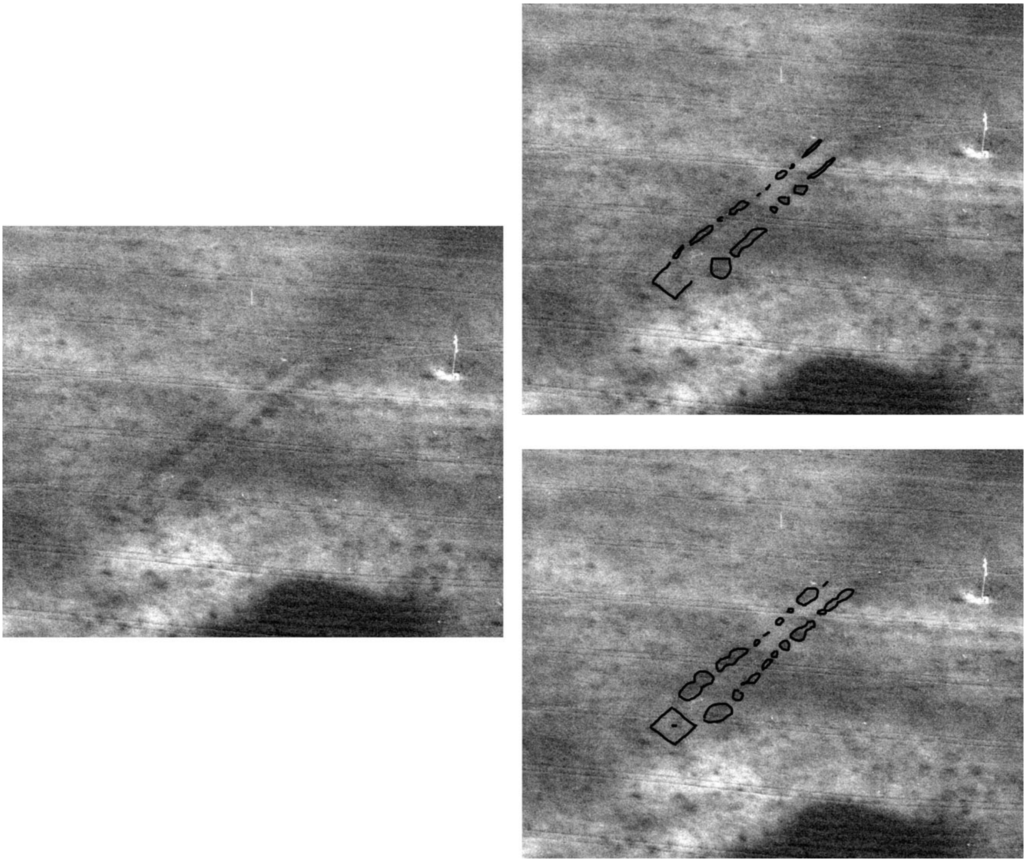

Figure 1. The cropmarks recorded on an aerial photograph (left) at Lagg in central Scotland illustrate the influence of experience and expectation on feature identification and interpretation. Without dwelling on the details of this example (see Brophy & Cowley 2005: 18–20) the mapping on the top right was driven by a desire to find Neolithic monuments, failed to consider the size of the monument, and was based on the interpreter's (D. Cowley) limited experience. The mapping on the bottom right was suggested by a colleague, and drew on a broader view of its context, less hide-bound by expectation and a desire to fill a hole in the distribution of Neolithic long barrows. This depiction would be interpreted as an unusual first millennium AD square barrow with an 'avenue' of pits, very different from the long barrow interpretation. Experienced (and reflective) interpreters will admit to such issues in their own work (e.g. Palmer 2011; Halliday 2013), and these show that an insistence on the primacy of the eye/brain combination for feature identification is unwise. Extract of A72255: Crown Copyright RCAHMS.

intelligible to be classed as anthropogenic in origin. Conversely, as automated processes identify everything that falls within the ruleset, there will be a high number of 'false positives' or features that fit the rules but are not archaeological (Trier & Pilø 2012).

Fundamentally, the two different approaches to image analysis should not be considered alternatives but rather complementary parts of an overall image assessment strategy. The idea of multi-scale analysis is not new to archaeology, but only in Norway have semi-automated detection techniques been incorporated as part of the strategic approach (Trier *et al.* 2009, Trier & Pilø 2012). It is our contention that without the appropriate incorporation of computer vision techniques we will fail to assess even a fraction of the data available to

Table 1. Details of two aerial prospection projects, summarised from Cowley *et al.* (2013).

	English Heritage National Mapping Programme (NMP)	Baden-Württemberg ALS Mapping, Germany
Aerial resource used	Predominantly aerial photographs, latterly some ALS	ALS project
Area covered	52 000km^2	35 000km^2
Personnel	Multiple operators for each project area	Single operator
Method	Manual classification and interpretation based on topographic, soil and vegetation contrast	Semi-automated classification of digital terrain models based on microtopographic variation (LRM), with manual verification
Timescale	20 years (since 1992/93)	6 years
Number of sites	Over 100 000 sites	600 000 sites (estimated)

us. Even for well-resourced nations with an established body of professionals trained in the field, current methods of interpretation are a major limitation on the potential exploitation of new imagery. For those without an infrastructure for image interpretation and cultural resource management the situation will prove even more acute.

Developing applications

Despite a steady increase in the sophistication and application of detection techniques in sectors as diverse as medical imaging, security, materials science and robotics, the techniques that have been tentatively used in archaeological contexts to date are frequently too simplistic to cope with the subtleties of archaeological feature detection. Lasaponara and Masini (2012) provide a very brief review of a handful of projects that have used computer detection techniques for classification, and while pointing out the inconclusive nature of the results, fail to observe that while the wholesale importation of techniques from the environmental sciences might be useful for monitoring the expression of known sites, it is inappropriate for the identification of archaeological remains, especially those that are previously undetected.

The over-simplicity and inflexibility of the modes of automated or semi-automated detection selected for analysis mean they are poorly suited to the diversity in shape, size, spectral and topographic properties of the features we categorise as archaeological (De Laet *et al.* 2007; Grøn *et al.* 2011: 2030). Embedded in this problem is the difficulty of precisely defining what an archaeological feature is. The truth of the matter is that regardless of the level of sophistication, a single detection algorithm could never adequately account for the range of feature types, let alone incorporate obfuscating factors such as plough damage, geological background variations and vegetation cover that significantly alter the appearance of a feature from a given normal descriptor. Picking apart these factors and providing a reasoned archaeological interpretation supported by observation and knowledge is a valuable specialist skill that computer detection methods, however sophisticated, will ultimately fail to replace.

This is why, far from being made redundant by automated techniques, skilled interpreters are vital to any computer vision process. The expert role is not simply to annotate training images or validate results, but as an integral part of the feature engineering process (see Domingos 2012 for discussion of the critical factors relating to the success of machine learning applications). To do this there has to be willingness to engage with both the practical and theoretical challenges of explaining and replicating human visual interpretation processes. With respect to openness of method, objectivity and repeatability of interpretations, asking 'what are we looking for?' and justifying the response can only be seen as a positive demonstration of critical thinking.

In our opinion the development of computer vision techniques for the analysis of airborne remote-sensing data is a high priority, and restructuring assessment methods and engagement in research are vital to providing assessment and protection of the historic environment. We identify the following areas for development, and although we recognise that some will require significant investment, we suggest them as the basis of a road-map to discipline-specific computer vision approaches that will pay dividends in the longer term.

1) A credible benchmark, based on Big Research

Large scale, high quality, systematic and discipline-aware research is required into the application of computer vision techniques to image analysis. This will need to incorporate robust and repeatable workflows that integrate computer vision and manual methods.

The creation of benchmark material for the evaluation of algorithms from an international range of archaeological and environmental contexts using a variety of data types is fundamental. Expert interpretations of the material can provide training and validation datasets, and support better understanding of why and how features are identified. Projects such as the PASCAL Object Recognition Database (PASCAL & PASCAL2 2013) provide exemplars for this type of work.

Key to the success of computer applications will be clearer definitions of archaeological features in various datasets. Although some work is already being undertaken for hyperspectral data under the DART project (Beck 2011), the knowledge-base needs to be expanded significantly. That can only practically be achieved via benchmarking and timely observations of the physical and chemical properties of features observed remotely. A clear distinction should be made between contrast detection and feature identification.

2) Improved protocols for recording pre-processing, data categorisation and metadata

Acceptance of data pre-processing as a beneficial first step for some archaeological analysis is critical to more mainstream adoption of semi-automated techniques. This includes masking high-contrast features (roads and buildings), high pass and low pass filters, local relief models to filter low relief and segmentation based on land cover. These techniques help to reduce large-scale variation in an image or model, and have been shown to help with the visualisation and detection of generally subtle archaeological traces. Metadata is needed to explain and qualify processing steps to improve the intelligibility of and confidence in the results of semi-automated techniques. To assist with standardisation, a categorisation system of different levels of processing (such as is common in environmental remote-sensing applications) should be established and included as metadata in all feature assessments. It would also be helpful to modify the language of feature detection to reflect the level of interpretation (i.e. potential archaeological feature or suspected archaeological anomaly).

Feature recording protocols should be expanded to include complementary environmental data such as land use, land management strategies, dates of feature visibility and proportion of known feature visible. If systematically recorded, this ancillary information will assist with the definition of archaeological features and help to explain gaps in identification.

3) Incorporation of multi-sensor data

The benefits of multi-sensor survey could be enhanced if computer vision techniques were used to combine different data from common geospatial locations, addressing a tendency to consider only one data source for computation, therefore limiting feature detection. The strength of computer-aided analysis of multi-sensor data has yet to be explored in any depth, and incorporation of tools such as decision tree analysis could begin to allow strategic targeting of attributes within multiple datasets that represent archaeological features. This is an area where computer vision techniques may have the most to contribute.

4) Innovative manual interpretation methods

Finally, in the interim, if manual detection is the best method we have to identify features of interest, some thought should be given to improving its rapidity and coverage. Some recent experimentation in the use of 'crowd-sourced' interpretations of ALS-derived visualisations is described by Duckers (2013). Information gathered could also include non-archaeological attributes like land use and 'noise' such as vegetation overlaying features and shadows, with the potential to derive a breadth of image information.

Acknowledgements

The authors would like to thank Carlo Tomasi, Chair of Computer Science, Duke University, for his comments on the draft and for expert input on the potential and practicalities of computer vision techniques, and to staff at English Heritage for providing estimates of manual aerial photograph interpretation rates. Numerous colleagues have contributed to the discussion of the issues at the heart of this paper, but particular thanks are due to Karsten Lambers, Rachel Opitz, Rog Palmer, Ole Risbøl and Øivind Due Trier.

References

BECK, A. 2011. Archaeological applications of multi/hyper-spectral data—challenges and potential, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 87–97. Budapest: Archaeolingua.

BENNETT, R., K. WELHAM, R.A. HILL & A. FORD. 2011. Making the most of airborne remote sensing techniques for archaeological survey and interpretation, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 99–106. Budapest: Archaeolingua.

– 2012. A comparison of visualization techniques for models created from airborne laser scanned data. *Archaeological Prospection* 19: 41–48. http://dx.doi.org/10.1002/arp.1414

BROPHY, K. & D. COWLEY. 2005. *From the air—understanding aerial archaeology*. Stroud: Tempus.

COWLEY, D.C. 2011. Remote sensing for European archaeology and heritage management—site discovery, interpretation and registration, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 43–55. Budapest: Archaeolingua.

COWLEY, D.C. & K. HULD SIGURÐARDÓTTIR. 2011. Remote sensing for archaeological heritage management, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 11–16. Budapest: Archaeolingua.

COWLEY, D.C., V. DE LAET & R.A. BENNETT. 2013. Auto-extraction techniques and cultural heritage databases, in W. Neubauer, I. Trinks, R. Salisbury & C. Einwögerer (ed.) *Proceedings of the 10th International Conference on Archaeological Prospection, Vienna, May 29–June 2 2013*: 406–408. Vienna: Ludwig Boltzmann Institute.

DOMINGOS, P. 2012. A few useful things to know about machine learning. *Communications of the ACM* 55(10): 78–87. http://dx.doi.org/10.1145/2347736.2347755

DUCKERS, G.L. 2013. Bridging the 'geospatial divide' in archaeology: community based interpretation of LIDAR data. *Internet Archaeology* 35. http://dx.doi.org/10.11141/ia.35.10

GOJDA, M. 2011. Remote sensing for the integrated study and management of sites and monuments—a Central European perspective and Czech case study, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 215–34. Budapest: Archaeolingua.

GRØN, O., S. PALMÉR, F. STYLEGAR, K. ESBENSEN, S. KUCHERYAVSKI & S. AASE. 2011. Interpretation of archaeological small-scale features in spectral images. *Journal of Archaeological Science* 38: 2024–30. http://dx.doi.org/10.1016/j.jas.2009.11.023

HALLIDAY, S. 2013. I walked, I saw, I surveyed, but what did I see? . . .and what did I survey? in R. Opitz & D.C. Cowley (ed.) *Interpreting archaeological topography: lasers, 3D data, observation, visualisation and applications*: 63–75. Oxford: Oxbow.

HANSON, W.S. 2010. The future of aerial archaeology in Europe. *Photo Interprétation: European Journal of Applied Remote Sensing* 46(1): 3–11.

HILL, R. 2009. *The roar of the butterflies*. London: HarperCollins.

HORNE, P. 2009. *A strategy for the National Mapping Programme*. Swindon: English Heritage.

DE LAET, V., E. PAULISSEN & M. WAELKENS. 2007. Methods for the extraction of archaeological features from very high-resolution Ikonos-2 remote sensing imagery, Hisar (southwest Turkey). *Journal of Archaeological Science* 34: 830–41. http://dx.doi.org/10.1016/j.jas.2006.09.013

LAMBERS, K. & I. ZINGMAN. 2013. Texture segmentation as a first step towards archaeological object detection in high resolution satellite images of the Silvretta Alps, in W. Neubauer, I. Trinks, R. Salisbury & C. Einwögerer (ed.) *Proceedings of the 10th International Conference on Archaeological Prospection, Vienna May 29–June 2 2013*: 327–29. Vienna: Ludwig Boltzmann Institute.

LASAPONARA, R. & N. MASINI. 2012. *Satellite remote sensing: a new tool for archaeology*. New York: Springer. http://dx.doi.org/10.1007/978-90-481-8801-7

PALMER, R. 2011. Knowledge-based aerial image interpretation, in D.C. Cowley (ed.) *Remote sensing for archaeological heritage management* (Europae Archaeologiae Consilium Occasional Papers 5): 283–91. Budapest: Archaeolingua.

PARCAK, S. 2009. *Satellite remote sensing for archaeology*. London: Routledge.

PASCAL & PASCAL2. 2013. Pattern analysis, statistical modelling and computational learning. Available at: http://pascallin.ecs.soton.ac.uk/ and http://pascallin2.ecs.soton.ac.uk/ (accessed 17 June 2014).

RISBØL, O., O.M. BOLLANDSÅS, A. NESBAKKEN, H. ØRKA, E. NÆSSET & T. GOBAKKEN. 2013. Interpreting cultural remains in airborne laser scanning generated digital terrain models: effects of size and shape on detection success rates. *Journal of Archaeological Science* 40: 4688–700. http://dx.doi.org/10.1016/j.jas.2013.07.002

SONKA, M., V. HLAVAC & R. BOYLE. 2008. *Image processing, analysis and machine vision*. Toronto: Thomson Learning.

TRIER, Ø. & L. PILØ. 2012. Automatic detection of pit structures in airborne laser scanning data. *Archaeological Prospection* 19: 103–21. http://dx.doi.org/10.1002/arp.1421

TRIER, Ø., S. LARSEN & R. SOLBERG. 2009. Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection* 16: 1–15. http://dx.doi.org/10.1002/arp.339

VERHAGEN, P. & L. DRĂGUŢ. 2012. Object-based landform delineation and classification from DEMs for archaeological predictive mapping. *Journal of Archaeological Science* 39: 698–703. http://dx.doi.org/10.1016/j.jas.2011.11.001

VERHOEVEN, G. 2012. Near-infrared aerial crop mark archaeology: from its historical use to current digital implementations. *Journal of Archaeological Method and Theory* 19: 132–60. http://dx.doi.org/10.1007/s10816-011-9104-5

WILSON, D.R. 2000. *Air photo interpretation for archaeologists*. London: Tempus.

Received: 10 September 2013; Accepted: 18 October 2013; Revised: 5 November 2013