ISPRS Journal of Photogrammetry and Remote Sensing 109 (2015) 165–177

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Wide-area mapping of small-scale features in agricultural landscapes using airborne remote sensing

Jerome O'Connell *, Ute Bradter, Tim G. Benton

School of Biology, Faculty of Biological Sciences, University of Leeds, LS2 9JT, UK

A R T I C L E I N F O

Article history:
Received 8 May 2015
Received in revised form 25 July 2015
Accepted 15 September 2015
Available online 9 October 2015

Keywords:
Random forest
Object orientated
Agriculture
Aerial photography
Spatial analysis
Classification

A B S T R A C T

Natural and semi-natural habitats in agricultural landscapes are likely to come under increasing pressure with the global population set to exceed 9 billion by 2050. These non-cropped habitats are primarily made up of trees, hedgerows and grassy margins and their amount, quality and spatial configuration can have strong implications for the delivery and sustainability of various ecosystem services. In this study high spatial resolution (0.5 m) colour infrared aerial photography (CIR) was used in object based image analysis for the classification of non-cropped habitat in a 10,029 ha area of southeast England. Three classification scenarios were devised using 4 and 9 class scenarios. The machine learning algorithm Random Forest (RF) was used to reduce the number of variables used for each classification scenario by 25.5 % ± 2.7%. Proportion of votes from the 4 class hierarchy was made available to the 9 class scenarios and where the highest ranked variables in all cases. This approach allowed for misclassified parent objects to be correctly classified at a lower level. A single object hierarchy with 4 class proportion of votes produced the best result (kappa 0.909). Validation of the optimum training sample size in RF showed no significant difference between mean internal out-of-bag error and external validation. As an example of the utility of this data, we assessed habitat suitability for a declining farmland bird, the yellowhammer (*Emberiza citronella*), which requires hedgerows associated with grassy margins. We found that ~22% of hedgerows were within 200 m of margins with an area >183.31 m^2. The results from this analysis can form a key information source at the environmental and policy level in landscape optimisation for food production and ecosystem service sustainability.

1. Introduction

Agricultural land covers approx 38% of the earth's terrestrial surface (FAO, 2014) and therefore plays a key role in biodiversity, conservation and ecosystem service delivery at a variety of spatial scales (Billeter et al., 2008; Tscharntke et al., 2005). However increasing pressures are likely on the fragmented habitats within these landscapes with the global population set to exceed 9 billion by 2050 driving demand that may require a doubling of food production (Godfray et al., 2010; Tilman et al., 2011). In agricultural landscapes non-cropped habitats are primarily made of features such as trees, hedgerows and grassy margins. For the purpose of this study a margin is a buffer strip ⩾2 m wide and a hedgerow is defined as a length of small trees and shrubs ⩾20 m long and ⩽5 m wide (Maddock, 2008). Trees are defined as having an individual crown >6 m^2 and may occur within a hedgerow, in isolation or as part of a woodland/forest. In the UK the total length of hedgerows fell from an estimated 800,000 km in 1956 to under 500,000 km in 1994 (Cornulier et al., 2011) to 477,000 km in 2007 (Carey et al., 2008). Recent reforms to the Common Agricultural Policy (CAP) have encouraged farmers to manage such features by means of financial payments through various agri-environmental schemes overseen by the Department for Environment, Food and Rural Affairs (DEFRA) in England. Features such as hedgerows and margins, are protected under UK (DEFRA, 1997, 2004) and EU (EU, 2007) law due to their importance as an ecological network across mono-cultured landscapes with distribution and connectivity having significant effects on landscape scale biodiversity and regional biota (Benton et al., 2003; Power, 2010). Many ecosystem services depend on the amount, quality and configuration of non-cropped land, as well as the landuse within fields (Benton, 2007; Billeter et al., 2008; Power, 2010). For example, simplification of the landscape through increased field size and reduced natural vegetation cover, especially of grassland areas, has been shown to increase pest

* Corresponding author.
E-mail address: oconnell.jerome@gmail.com (J. O'Connell).

http://dx.doi.org/10.1016/j.isprsjprs.2015.09.007
0924-2716/

damage due to lower populations of natural enemies (Gardiner et al., 2009).

Given the importance of the amount, connectivity, heterogeneity and quality of non-cropped habitat for biodiversity and other ecosystem services, spatially extensive knowledge on the location and state of these habitats has now been recognized as a key variable in mapping and modeling ecosystem service delivery and sustainability at local, regional and national scales (Dale and Polasky, 2007; Kremen et al., 2007; Watson et al., 2011). Many of the processes involved need to be assessed at landscape level which makes traditional field based surveys expensive and time consuming.

Remote sensing has long since been used to map biodiversity at a variety of spatial scales (Turner et al., 2003). However non-cropped features in the UK are often below the spatial resolution of many satellite sensors (i.e. <5 m), therefore alternative platforms are required for accurate delineation of their extent (O'Connell et al., 2013a). Some studies have used image fusion to enhance the spatial resolution of multispectral bands (Aksoy et al., 2010) while others have looked at sub-pixel image classification for the detection of small scale woody elements in the landscape (Foschi and Smith, 1997; Thornton et al., 2007). Many other approaches focus on the use of edge detection kernels to detect spectral boundaries which may be indicative of hedgerows or margins in agricultural landscapes (Fauvel et al., 2012; Rydberg and Borgefors, 2001). All these approaches are pixel based and rely on either high spatial resolution panchromatic data for contrast or multispectral data to classify the features based on spectral response. Other approaches have classified trees, hedge and shrub vegetation by combining multispectral and structural data via stereo imaging (Tansey et al., 2009), Light Detection And Ranging (LiDAR) (Hellesen and Matikainen, 2013) and radar (Scholefield et al., 2012). While these approaches offer robust mapping of tall vegetation, they generally don't enhance the classification of surface vegetation such as grassy margins, can be costly to acquire and generally are not suitable for regional scale mapping due to low spatial resolution or low spatial coverage.

An alternative approach to the use of pixels is the use of objects which can add additional information to features of interest which can then be utilised in an Object Based Image Analysis (OBIA) protocol (Blaschke, 2010). OBIA uses a variety of spectral, textural, geometric, thematic and contextual attributes built from the aggregation of homogeneous pixels into real world objects, therefore the size of uncorrelated feature space is significantly increased when compared to traditional pixel based approached (Benz et al., 2004; Mallinis et al., 2008; Myint et al., 2011). Non-cropped features in structured agricultural landscapes generally have high geometric and contextual properties; e.g. margins typically have a high length-to-width ratio, show a high contrast to neighbouring features such as hedges and are located at the edge of fields. Several studies have used OBIA in the classification of non-cropped features such as trees and hedges with varying levels of success (Bock et al., 2005; Mueller et al., 2004; Sheeren et al., 2009; Tansey et al., 2009; Vannier and Hubert-Moy, 2008). Classification in OBIA has been dominated by algorithms such as maximum likelihood, nearest neighbour and Knowledge Based Classifiers (KBC) (Blaschke, 2010). A KBC incorporates expert knowledge in building a set of rules that utilise the attributes of each object in the image and have proved successful in the classification of non-cropped areas in the UK (O'Connell et al., 2013a; Tansey et al., 2009). However the creation of a robust KBC is an iterative process which can be time consuming with respect to the selection of suitable features and thresholds (Stumpf and Kerle, 2011). Model transferability can be a significant issue where the thresholds and membership functions within the rule-base break down when the KBC is applied to data which was taken at a different time or location (O'Connell et al., 2013a). An alternative approach which is gaining popularity within the remote sensing community is the use of machine learning or ensemble algorithms such as Support Vector Machine (SVM) and Random Forest (RF) in OBIA (Duro et al., 2012). Such algorithms have advantages over conventional algorithms based on their ability to detect subtle and complex patterns in high dimensional data using robust statistical techniques with a high degree automation (Blaschke, 2010; Rodriguez-Galiano et al., 2012). Previous experience of RF (Breiman, 2001) from some of the authors of this study in the classification of complex habitats in the Yorkshire Dales National Park in the UK (Bradter et al., 2011) gave an indication of its potential when applied to environmental and remote sensing data.

RF is based on an ensemble of decision trees which are each grown on random selections of two thirds of the data with replacement. This "bagging" approach makes the algorithm more insensitive to noise in the data (Rodriguez-Galiano et al., 2012), including variations in reflectance due to solar zenith or Biodirectional Reflectance Distribution Function (BRDF) (Chan and Paelinckx, 2008). This concept of machine learning by randomisation over multiple iterations allows for discernible pattern to emerge from highly dimensional data. The set of variables used at each decision node is randomly selected which can reduce the strength of individual trees but also reduces correlation between trees and thus reduces the generalisation error (Liaw and Wiener, 2002). The proportion between misclassifications and the total number of Out Of Bag (OOB) elements (i.e. the remaining one third of data) contributes to an unbiased estimate of generalisation error. This error converges as the number of trees increases; therefore adding more trees does not over fit the data (Cutler et al., 2007). RF uses the "best" variables at each node based on node purity. Several options to calculate variable importance exist, including permutation importance which is calculated by randomly permutating all values in the selected variable and using the difference in OOB error as an indication of the importance of that variable to the classifier. Pruning of trees is not necessary as the final classification is based on the majority vote of all trees within the forest. The package randomForest 4.6–7 (Liaw and Wiener, 2002) was used in the R (3.0.1) statistical coding environment (R Development Core Team, 2014), which is based on the original Fortran programs by Breiman and Cutler (https://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm).

The objective of this study was to create a novel and robust classification protocol for the mapping of non-cropped features in a "case study" agricultural landscape. The protocol needed to be semi-automated to enable wide area mapping of such features with a minimal number of variables. Based on these criteria and results from a previous studies (Bradter et al., 2013; O'Connell et al., 2013a) as well as a review of some comparative studies with other algorithms (Chan and Paelinckx, 2008; Duro et al., 2012; Lawrence et al., 2006; Pal, 2005; Rodriguez-Galiano et al., 2012), it was felt that OBIA with the ensemble classifier RF may yield best results. The combination of both approaches in remote sensing is a relatively new area of research and yields some uncertainties in the areas of optimisation of RF parameters, model transparency, training and validation size, hierarchical accuracy assessment and feature selection/importance with respect to class and object hierarchy. This study addresses some of the aforementioned uncertainties through the mapping of non-cropped areas using OBIA and the machine learning classifier RF.

2. Materials

2.1. Study area

The study area was located in East Anglia, England (52°19′07″ N, 0°49′43″ E) in an intensively managed agricultural landscape of