Identification of linear vegetation elements in a rural landscape using LiDAR point clouds

Chris Lucas[a,b,c], W Bouten[a], Zsofia Koma[a], W. Daniel Kissling[a], Arie C. Seijmonsbergen[a,*]

[a] *Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands*
[b] *Spatial Information Laboratory (SPINLab), Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam*
[c] *Geodan, President Kennedylaan 1, 1079 MB Amsterdam, The Netherlands*

Abstract

Modernization of agricultural land use across Europe is responsible for a substantial decline of linear vegetation elements such as tree lines, hedgerows, riparian vegetation and green lanes. These linear objects have an important function for biodiversity, e.g. as ecological corridors and local habitats for many animal and plant species. Knowledge on their spatial distribution is therefore essential to support conservation strategies and regional planning in rural landscapes, but detailed inventories of such linear objects are often lacking. Here, we propose a method to detect linear vegetation elements in agricultural landscapes using classification and segmentation of high resolution LiDAR point data. To quantify the 3D structure of vegetation we applied point cloud analysis to identify features based on the local neighborhood. As a preprocessing step, we removed planar surfaces, such as grassland, bare soil and water bodies, from the point cloud using scatter information. We then applied a random forest classifier to separate the remaining points into 'vegetation' and 'other'. Subsequently, a region growing algorithm allowed to segment 2D rectangular objects, which were then classified into linear objects based on their elongatedness. We evaluated the accuracy of the linear objects against a manually delineated set. This assessment showed that the majority of vegetation objects were correctly identified. These results are a promising first step for testing our method in other regions and for upscaling it to broad spatial extents. This would allow producing detailed inventories of linear vegetation elements at regional and continental scales in support of biodiversity conservation and regional planning in agricultural and other rural landscapes.

Keywords: Linear vegetation, LiDAR, Point cloud, Object recognition, Classification, Segmentation, Rectangularity, Agricultural landscapes

1. Introduction

The European landscape has dramatically changed during the Holocene as a result of human impact and climatic change (Turner, 1989; Marquer et al., 2017). Especially since the industrial revolution, landscapes

*Corresponding author. E-mail address: A.C.Seijmonsbergen@uva.nl (A. C. Seijmonsbergen).

have been deforested and reshaped into rural and agricultural landscapes. These are dominated by a mosaic of grasslands, forests and urban areas, separated or connected by linear landscape objects such as roads, ditches, tree lines, lynchets and hedgerows (Bailly et al., 2008; Meyer et al., 2012; Van der Zanden et al., 2013). The distribution, abundance and richness of species in these landscapes is related to the amount, height, length and quality of linear vegetation objects (Aguirre-Gutiérrez et al., 2016; Spellerberg and Sawyer, 1999; Croxton et al., 2005). The same holds true for the dispersal of seeds and the flow of matter, nutrients and water (Turner, 1989; Burel, 1996). Additionally, linear infrastructures such as roads and railways form barriers which lead to habitat fragmentation. In contrast, green lanes which are flanked by hedges and/or tree lines may form connecting corridors. Hence, linear vegetation elements are of key importance for biodiversity in agricultural landscapes. Furthermore, the importance of linear vegetation objects lies also in the fact that they provide essential services on a landscape scale and as such are not merely an artifact of historical management systems (Scholefield et al., 2016). A wider audience has become aware that historic agricultural practices are part of the cultural heritage (Jongman, 2004; Gobster et al., 2007) and need to be conserved. However, the occurrence of green lanes and hedgerows has strongly diminished in many countries (Boutin et al., 2001; Stoate et al., 2001). This is mostly a consequence of larger agricultural fields, monocultures and a reduction in non-crop features which reduces the complexity and diversity of landscape structure (Croxton et al., 2005). Detailed knowledge of the spatial occurrence, current status, frequency and ecological functions of linear vegetation objects in a landscape is therefore of key importance for biodiversity conservation and regional planning.

The mapping of linear vegetation structures has traditionally been done with visual interpretations of aerial photographs in combination with intensive field campaigns (Aksoy et al., 2010). However, this approach is time-consuming and has limited transferability to larger areas. New methods have therefore been developed that use raster images to map linear vegetation objects by using their spectral properties in visible or infrared wavelengths, e.g. from SPOT, ASTER and Landsat imagery (Thornton et al., 2006; Vannier and Hubert-Moy, 2014; Tansey et al., 2009). This allows an automated and hierarchical feature extraction from very high resolution imagery (Aksoy et al., 2010). Despite these developments, comprehensive high-resolution inventories of linear vegetation elements such as hedgerows and tree lines are lacking at regional and continental scales. The lack of such high-resolution measurements of 3D ecosystem structure across broad spatial extents impedes major advancements in animal ecology and biodiversity science (Kissling et al., 2017). On a European scale, density maps of linear landscape elements have been produced at 1 km^2 resolution through spatial modelling of 200,000 ground observations (Van der Zanden et al., 2013). However, these maps strongly depend on spatial interpolation methods as well as regional environmental and socio-economic variation and therefore contain a considerable amount of uncertainty in the exact distribution of linear landscape elements. High-resolution measurements of 2D and 3D ecosystem structures derived from cross-national remote sensing datasets are therefore needed to quantify linear woody objects across broad

spatial extents (Kissling et al., 2017).

An exciting development for quantifying 3D ecosystem structures is the increasing availability of high-resolution remote sensing data derived from Light Detection and Ranging (LiDAR) (Lim et al., 2003). LiDAR data have important properties which are useful for the detection, delineation and 3D characterization of vegetation, such as their physical dimensions x, y, z, laser return intensity, and multiple return information (Lefsky et al., 2002; Eitel et al., 2016). Vegetation partly reflects the LiDAR signal and usually generates multiple returns, including a first return at the top of the canopy and a last return on the underlying terrain surface. This provides valuable information for separating vegetation from non-vegetation (Lim et al., 2003). Moreover, the intensity values describe the strength of the returning light, which depends on the type of surface on which it is reflected and therefore provides key information on the surface composition (Song et al., 2002). The shape and internal structure of vegetation can be analyzed by classifying information from the different return values and a variety of features, which can be calculated from the point cloud (Lim et al., 2003; Weinmann et al., 2015). Some applications of using airborne LiDAR data to quantify linear landscape elements already exist, e.g. the extraction of ditches in a Mediterranean vineyard landscape (Bailly et al., 2008). However, the characterization of linear vegetation elements in rural and agricultural landscapes from LiDAR point clouds is mostly lacking. Nevertheless, the increasing availability of nation-wide and freely accessible LiDAR data in several European countries provides exciting new avenues for characterizing 3D vegetation structures in agricultural landscapes (Kissling et al., 2017).

Here, we present a transparent and accurate method for classifying linear vegetation objects from LiDAR point clouds in an agricultural landscape. We develop the method using free and open source data and analysis tools and apply it for characterizing various linear vegetation objects in a rural landscape of the Netherlands containing agricultural fields, grasslands, bare soil, roads and buildings. While the identification of linear objects (e.g. the automated delineation of roads) is often based on raster-based remotely sensed imagery (Quackenbush, 2004), the detection of linear vegetation objects is more complex due to their 3-dimensional shape, size and variety. We therefore use a method which allows us to directly classify the point cloud using machine learning algorithms (Yan et al., 2015). Using fourteen features based on echo and local geometry information of the LiDAR point cloud, we apply a machine learning algorithm to classify the vegetation points in the point cloud. We then use a region growing algorithm to segment rectangular objects, and apply elongatedness as a criterion to classify linear objects. The accuracy of the method is tested against manually annotated datasets based on high resolution orthophotos and field survey. Our method provides a promising first step for upscaling the detection of linear vegetation objects in agricultural landscapes to broad spatial extents.

2. Data and study area

2.1. LiDAR and orthophoto data

Raw LiDAR point cloud data were retrieved from "Publieke Dienstverlening op de Kaart" (PDOK), an open geo-information service of the Dutch government.[1] The data are part of the "Actueel Hoogtebestand Nederland 3" (AHN3) dataset, which is collected between 2014 and 2019. The point density is on average over 10 points/m^2 and includes multiple discrete return values as well as intensity data. The dataset is collected in the first quarter of each year when deciduous vegetation is leafless (AHN, 2016). Nevertheless, the return signal is sufficiently strong to retrieve a useful scan of the vegetation cover. Freely available very high resolution (VHR) true color orthophotos from PDOK with a resolution of 25cm were consulted for validation purposes.[2]

All data were analyzed using free and open source software.[3] The scripting was performed in Python (3.6.4) using the NumPy (1.13.3) (Walt et al., 2011), SciPy (1.0.0) (Jones et al., 2001), pandas (0.20.3) (McKinney et al., 2010), and scikit-learn (0.19.1) (Pedregosa et al., 2011) libraries. For the preprocessing of data, the software LASTools (version 160429) and CloudCompare (v2.10beta) were used. CloudCompare was also used for visualizing and down-sampling the point cloud.

2.2. Study area

The case study area is located in a rural landscape in the center of the Netherlands south of Amsterdam (figure 1). The area is about 1.6 km from east to west and 1.2 km from north to south, spanning an area of almost 2 million square meters. The point density of the point cloud in the area is 22.49 points/m^2. The area contains numerous woody linear objects of varying geometry, ranging from completely straight to curved, isolated or connected to other linear or non-linear objects. Examples of vegetation and non-vegetation elements are planted forest patches, hedges, green lanes, isolated farms, ditches, a river, dykes and a road network (figure 1). This heterogeneity within a small area ensured that both the classification of vegetation and delineation of linear objects can be efficiently trained and tested.

3. Method

An overview of the method is presented in figure 2. There are three main routines: feature extraction, classification of vegetation, and segmentation of linear objects. In the feature extraction we use the data of the point and its neighbors to get additional information that can help determine to what class a point belongs to. We then use this information to classify which points belong to higher than herb vegetation. These points we subsequently segment into rectangular region, which can be checked for linearity. Details of the various steps which will be described in the following sections.

[1] https://www.pdok.nl/nl/ahn3-downloads
[2] https://www.pdok.nl/nl/service/wms-luchtfoto-beeldmateriaal-pdok-25-cm-rgb
[3] https://github.com/clucas111/delineating-linear-elements

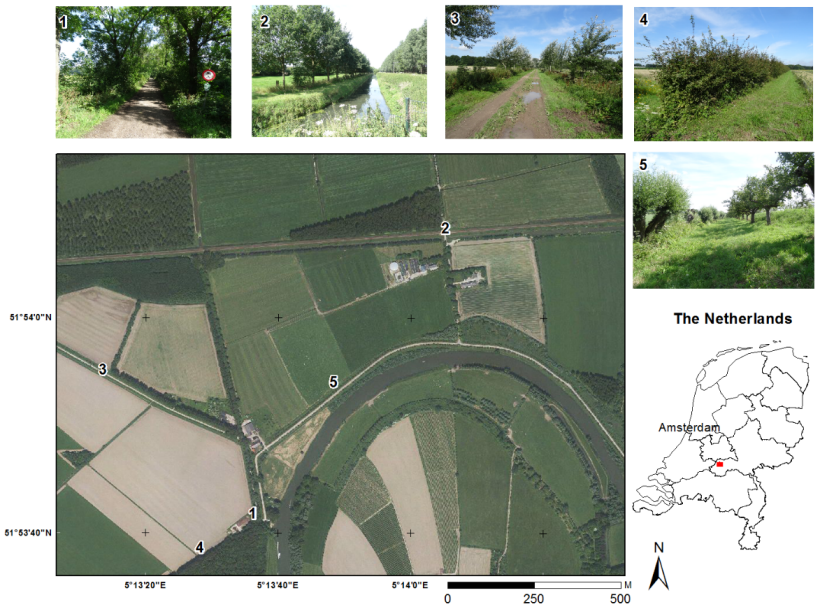

Figure 1: Location of the study area in the central part of The Netherlands. The true color aerial photo (PDOK) shows several linear objects in the rural landscape related to agricultural fields, grasslands, bare soil and infrastructure such as (un)paved roads and farmhouses. The numbered photo insets show a selection of the variety of linear vegetation elements, such as green lanes (photo 1), planted high tree lines along ditches, low and high shrubs/copse (photo 3), hedges (photo 4) and isolated traditional fruit trees.

3.1. Feature extraction

The relevance of the various input features has extensively been studied to separate urban from vegetation objects (Chehata et al., 2009; Guo et al., 2011; Mallet et al., 2011), but concise information for vegetation classification is scarce. After reviewing relevant literature, we selected fourteen features (table 1). These features are based on information from echo and local neighborhood information (geometric and eigenvalue based) and their qualities are considered to be efficient for discriminating vegetation objects from point clouds (Chehata et al., 2009).

3.1.1. Point-based features

The point cloud $\mathcal{P}$ is a set of points $\{p_1, p_2, \ldots, p_n\} \in \mathbb{R}^3$, where each point p_i has x, y and z coordinates. In addition, an intensity value (I), a return number (R), and a number of returns (R_t) of the returned signal are stored. Since we do not have all the data required to do a radiometric correction of the intensity data we omitted this feature for the classification (Kashani et al., 2015). As an additional echo-based feature we calculated a normalized return number (R_n) (table 1).

3.1.2. Neighborhood-based features

In addition to point-based features, we calculated four geometric features related to local neighborhoods. We defined a neighborhood set $\mathcal{N}_i$ of points $\{q_1, q_2, \ldots, q_k\}$ for each point p_i, where $q_1 = p_i$, by using the k-nearest neighbors method with $k = 10$ points. This method performs well for datasets that vary in

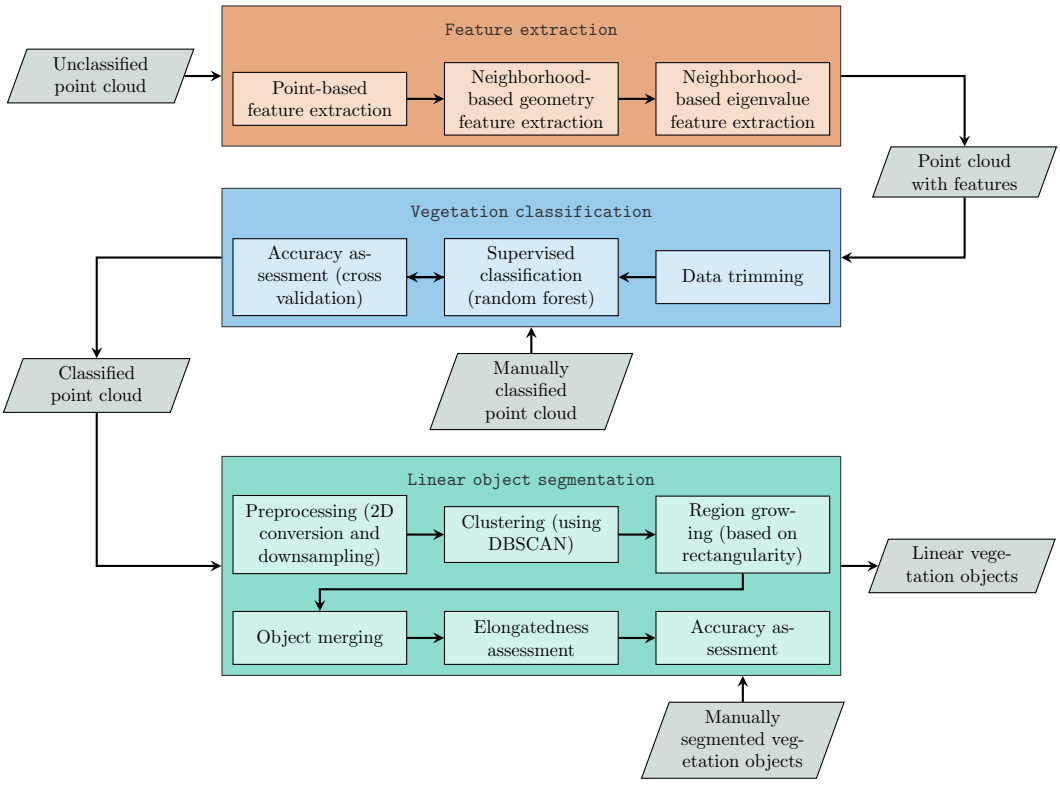

Figure 2: Workflow for feature extraction (orange), classification (blue), and segmentation of linear objects (green), with datasets represented as parallelograms and processes as rectangles.

point densities (Weinmann et al., 2014). The four geometric features were: the height difference, the height standard deviation, the local radius and local point density (table 1).

We further calculated eigenvalue-based features which are used to describe the distribution of points in a neighborhood (Hoppe et al., 1992; Chehata et al., 2009). We used the local structure tensor to estimate the surface normal and to define surface variation (Pauly et al., 2002). The structure tensor describes the dominant directions of the neighborhood of a point by determining the covariance matrix of the x, y and z coordinates of the set of neighborhood points and computing the eigenvalues ($\lambda_1, \lambda_2, \lambda_3$, where $\lambda_1 > \lambda_2 > \lambda_3$) of this matrix and ranking them based on the eigenvalue values. Hence, the magnitude of the eigenvalues of this covariance matrix describe the spread of points in the direction of the eigenvector. The points are linearly distributed if the eigenvalue of the first principle direction is significantly bigger than the other two ($\lambda_1 \gg \lambda_2 \approx \lambda_3$), planarly distributed if the eigenvalues of the first two principle directions are about equal and significantly larger than the third ($\lambda_1 \approx \lambda_2 \gg \lambda_3$), and the points are scattered in all directions if all eigenvalues are about equal ($\lambda_1 \approx \lambda_2 \approx \lambda_3$). These and additional properties are quantified using formulas (table 1). The eigenvector belonging to the third eigenvalue is equal to the normal vector ($\vec{N} = (N_x, N_y, N_z)$) (Pauly et al., 2002).

Table 1: The features used for classification, split into two main groups: point-based and neighborhood-based. The point-based features are based on echo information and the neighborhood-based features are based on the local geometry and eigenvalue characteristics.

Feature group	Feature	Symbol	Formula	Reference
Point				
- Echo	Number of returns	R_t		
	Normalized return number	R_n	R/R_t	Guo et al. (2011)
Neighborhood				
- Geometric	Height difference	Δ_z	$\max_{j:\mathcal{N}_i}(q_{z_j}) - \min_{j:\mathcal{N}_i}(q_{z_j})$	Weinmann et al. (2015)
	Height standard deviation	σ_z	$\sqrt{\frac{1}{k}\sum_{j=1}^{k}(q_{z_j} - \overline{q_z})^2}$	Weinmann et al. (2015)
	Local radius	r_l	$\max_{j:\mathcal{N}_i}(\lvert p_i - q_j \rvert)$	Weinmann et al. (2015)
	Local point density	D	$k/(\frac{4}{3}\pi r_l^3)$	Weinmann et al. (2015)
- Eigenvalue	Normal vector Z	N_z		Pauly et al. (2002)
	Linearity	L_λ	$\frac{\lambda_1-\lambda_2}{\lambda_1}$	West et al. (2004)
	Planarity	P_λ	$\frac{\lambda_2-\lambda_3}{\lambda_1}$	West et al. (2004)
	Scatter	S_λ	$\frac{\lambda_3}{\lambda_1}$	West et al. (2004)
	Omnivariance	O_λ	$\sqrt[3]{\lambda_1\lambda_2\lambda_3}$	West et al. (2004)
	Eigenentropy	E_λ	$-\lambda_1\ln(\lambda_1) - \lambda_2\ln(\lambda_2) - \lambda_3\ln(\lambda_3)$	West et al. (2004)
	Sum of eigenvalues	$\sum_\lambda$	$\lambda_1 + \lambda_2 + \lambda_3$	Mallet et al. (2011)
	Curvature	C_λ	$\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$	Pauly et al. (2002)

3.2. Vegetation classification

The fourteen features served as input for the classification of vegetation, which required data trimming, a supervised vegetation classifier and accuracy assessment.

3.2.1. Data trimming

To facilitate efficient processing, points that certainly do not belong to higher than herb vegetation were removed from the dataset. These points are characterized by a locally planar neighborhood and are selected on the basis of sphericity feature (table 1). In this way, points with a scatter value $S_\lambda < 0.05$ were removed. This threshold was conservative to ensure a large reduction of the data size, while still preserving the higher than herb vegetation points.

3.2.2. Supervised classification

A random forest classifier provides a good trade-off between classification accuracy and computational efficiency (Breiman, 2001; Weinmann et al., 2015). The random forest algorithm creates a collection of

decision trees, where each tree is based on a random subset of the training data (Ho, 1998). Random forest parameters such as the maximum number of features, minimal samples per leaf, minimal samples per split and the ratio between minority and majority samples were optimized using a cross validated grid search. During the grid search a range of applicable values were chosen for each parameter and all combinations were tested and evaluated for performance using cross validation. To save time first a course grid was created to identify the region of best performance and subsequently a finer grid was made to find the best performing parameter set in that region (Hsu et al., 2003).

The trimmed point cloud is imbalanced, and includes a lot more vegetation than 'other' points. Imbalanced training data can lead to undesirable classification results (He and Garcia, 2009). Therefore, we used a balanced random forest algorithm. In this algorithm the subsets are created by taking a bootstrap sample from the minority class and a random sample from the majority class with the same sample size as the minority class sample (Chen et al., 2004). By employing enough trees eventually all majority class data are used, while still maintaining a balance between the two classes. The decision trees were created using a Classification and Regression Tree (CART) algorithm (Breiman et al., 1984). A manual annotation of the trimmed point cloud into 'vegetation' and 'other' classes was done using an interpretation of the point cloud and high resolution aerial photos.

3.2.3. Accuracy assessment

The receiver operating characteristic (ROC) curve (Bradley, 1997), the Matthews correlation coefficient (MCC) (Matthews, 1975) and the geometric mean (Kubat et al., 1998) were used in the accuracy assessment. This allowed to have good performance metrics even when dealing with an imbalanced dataset (Kohavi et al., 1995; Sun et al., 2009; López et al., 2013). To create a ROC curve, the true positive (TP) rate is plotted against the false positive (FP) rate at various decision thresholds. The area under a ROC curve (AUROCC) is a measure for the performance of the classifier (Bradley, 1997). The MCC analyzes the correlation between the observed and the predicted data and is defined as:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{1}$$

where TN are the true negatives and FN the false negatives. The geometric mean is defined as:

$$\text{Geometric mean} = \sqrt{\text{producer's accuracy class 1} \times \text{producer's accuracy class 2}} \tag{2}$$

The MCC, AUROCC and the geometric mean were obtained using a 10-fold cross validation. This is done by splitting the data into 10 randomly mutually exclusive subsets and using a subset as testing data and a classifier trained on the remaining data (Kohavi et al., 1995).

Figure 3: The vegetation points of a piece of tree line within the research area before (blue) and after (red) downsampling the point cloud, plotted on top of the high resolution orthophoto, in RD coordinates.

3.3. Linear object segmentation

To segment the vegetation points into linear regions we did a preprocessing step, clustered the points using a DBSCAN, applied a region growing algorithm, merged nearby and aligned objects, evaluated their elongatedness and assessed the accuracy.

3.3.1. Preprocessing

The point cloud was converted to 2D by removing the z-coordinate of the vegetation points. In addition, the data were spatially downsampled to 1 meter distance between vegetation points. This substantially decreased computation time without losing too much precision (figure 3).

3.3.2. Clustering

After reducing the amount of points we clustered the remaining points together using a DBSCAN clustering algorithm (Ester et al., 1996). This algorithm is able to quickly cluster points together based on density and removes outlying points in the process. This decreases the processing time needed in the subsequent region growing step, since the amount of possible neighboring points is reduced.

3.3.3. Region growing

Region growing is an accepted way of decomposing point clouds (Rabbani et al., 2006; Vosselman, 2013) or raster imagery (Blaschke et al., 2014) into homogeneous objects. Normally, seed locations are selected

and regions are grown based on proximity and similarity of the attributes of the points. In our method, seed selection was based on the coordinates and regions were grown based on proximity and a rectangularity constraint. The proximity is constrained by taking the eight nearest neighbors of each point during the growing process. The rectangularity of an object is described as the ratio between the area of an object and the area of its minimum bounding rectangle (Rosin, 1999). The minimum bounding rectangle (figure 4b) is computed with rotating calipers (Toussaint, 1983). First a convex hull is constructed (figure 4a) by using the QuickHull algorithm (Preparata and Shamos, 1985). The minimum bounding rectangle has a side collinear with one of the edges of the convex hull (Freeman and Shapira, 1975). Hence the minimum bounding rectangle can be found by rotating the system by the angle the edges of the convex hull make with the x-axis and checking the bounding rectangles of each rotation. The area of the object can be calculated by computing the concave hull of the set of points belonging to the object (figure 4c). For this we use a concave hull algorithm using a k-nearest neighbors approach (Moreira and Santos, 2007). This algorithm starts with finding a point with a minimum or maximum of one of the coordinates. Subsequently it goes around the points by repeatably finding the neighbor which makes an edge that makes the largest counterclockwise angle with the previous edge. This process is continued until the current point is the starting point. Finally a check is done to ensure all the points fall within the hull. If this is not the case the k is increased by 1 and the algorithm repeated until an acceptable concave hull has been found.

In this way, for each cluster, points with the minimum x-coordinate and its 10 closest neighbors were used as the starting region. Subsequent points were added as long as the regions rectangularity did not drop below a set threshold (figure 5). A sensitivity analysis of this threshold value on a subset of the data showed the best performance of the algorithm when this value was between 0.5 and 0.6, with marginal differences in performance inbetween these values. Therefore the threshold value was set at 0.55. After a region is grown, the growing procedure was repeated for the next region until the entire cluster is segmented into rectangular regions.

3.3.4. Object merging

The resulting objects can be fragmented as the result of minor curves in the linear vegetation or small interruptions in vegetation. These objects were merged when they were in close proximity, faced a similar compass direction, and were aligned. The compass direction was determined by computing the angle between one of the long sides of the minimum bounding box and the x-axis. The alignment was checked by comparing the angle of the line between the two center points with the directions of the objects. Once merged the lengths of the objects were added and the maximum of the widths taken as the new width.

3.3.5. Elongatedness

The merged objects were assessed for linearity by resolving the elongatedness of an object, which is defined as the ratio between its length and its width (Nagao and Matsuyama, 2013). The definition of a

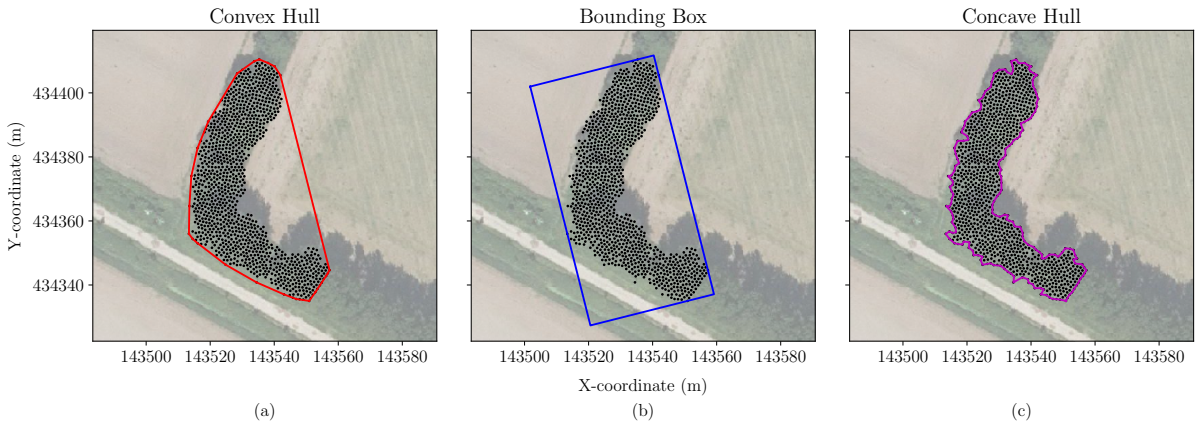

Figure 4: The downsampled vegetation points of a piece of tree line within the research area plotted on top of the high resolution orthophoto in RD coordinates, showing the different hulls used during the region growing algorithm: the convex hull (a), the minimal oriented bounding box (b), and the concave hull (c). During the region growing the rectangularity is calculated by dividing the area of the concave hull (c) by the area of the bounding box (b).

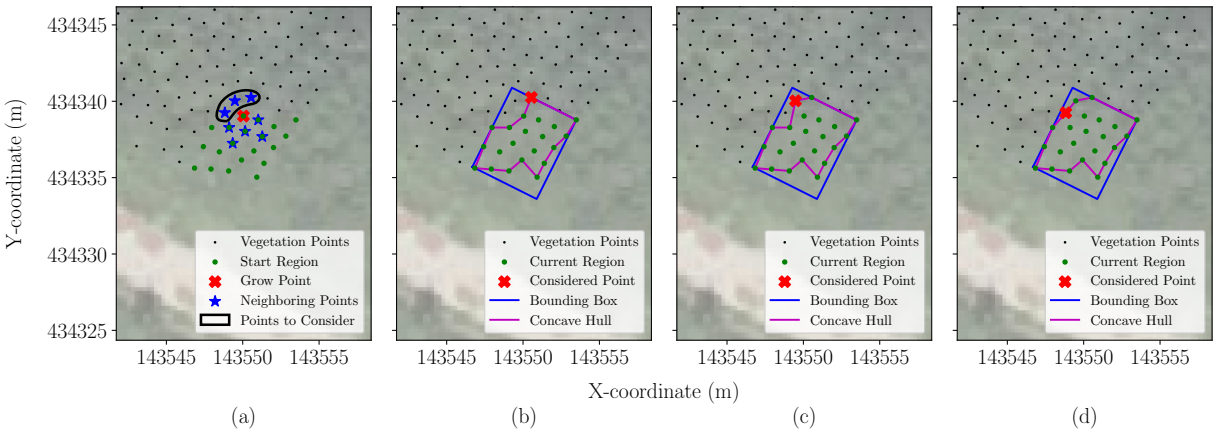

Figure 5: An example of the region growing process for one point. First the eight nearest neighbors are computed and the neighbors which are not already part of the region are considered to be added to the region (a). A bounding box and concave hull is computed for the region with the considered point added and the rectangularity is calculated (b). If this rectangularity is above a certain threshold the point is added to the region. Subsequently the process is repeated for the other points to consider (c, d). When all points to consider are checked, a next point of the region is check for nearest neighbors and the whole process is repeated until all points of the region, including the ones which are added during the growing process, have been checked once.

Table 2: A confusion matrix showing the predicted classes against the actual classes of the points. These are accumulated over the 10 fold cross-validation.

		Predicted	
		Vegetation	Other
Actual	Vegetation	974177	22908
	Other	8171	47999

linear object is not clearly defined and consequently somewhat arbitrary. After analysing the results using different values we set the minimum elongatedness at 2.5 and a maximum width of 60 meters.

3.3.6. Accuracy assessment

The accuracy of the delineated linear objects was assessed by calculating the users, producers and overall accuracy, as well as the harmonic mean of the precision and recall (F1), kappa and MCC scores (Congalton and Green, 2008). We manually annotated the vegetation data into linear and nonlinear objects, after converting the classified vegetation points into polygons. Consequently this assessment evaluates the accuracy of the segmentation given the accuracy of the vegetation points. Differencing of the automated and manually constructed data resulted in confusion matrices to compare true positives, true negatives, false positives and false negatives in the area.

4. Results

4.1. Vegetation classification

The vegetation classification resulted in a map showing the different classes (figure 6). The data points which were removed during the preprocessing are displayed in grey. This class contains data points that certainly do not contain higher than herb vegetation, as it mainly contains bare soil, grassland and water bodies (compare figure 1). The blue class represents data points which were not removed during the preprocessing, but were classified as 'other' during the supervised classification (mainly building edges, ditches and railroad infrastructure). The green class represents the classified vegetation. The accuracy of this classification is presented in a confusion matrix of predicted versus actual classes (table 2). From the confusion matrix a producers accuracy of 0.98 for 'vegetation' and of 0.85 for 'other' was calculated. The AUROCC of 0.98 shows that 'vegetation' and 'other' class are separated well, which is supported by the MCC of 0.76, indicative of a positive correlation between the predicted and observed classes, and the geometric mean of 0.90.

4.2. Linear object segmentation

The vegetation was segmented using our region growing algorithm and the resulting objects were filtered for linearity. We compare the result with our manual segmentation (figure ..). The areas that were correctly

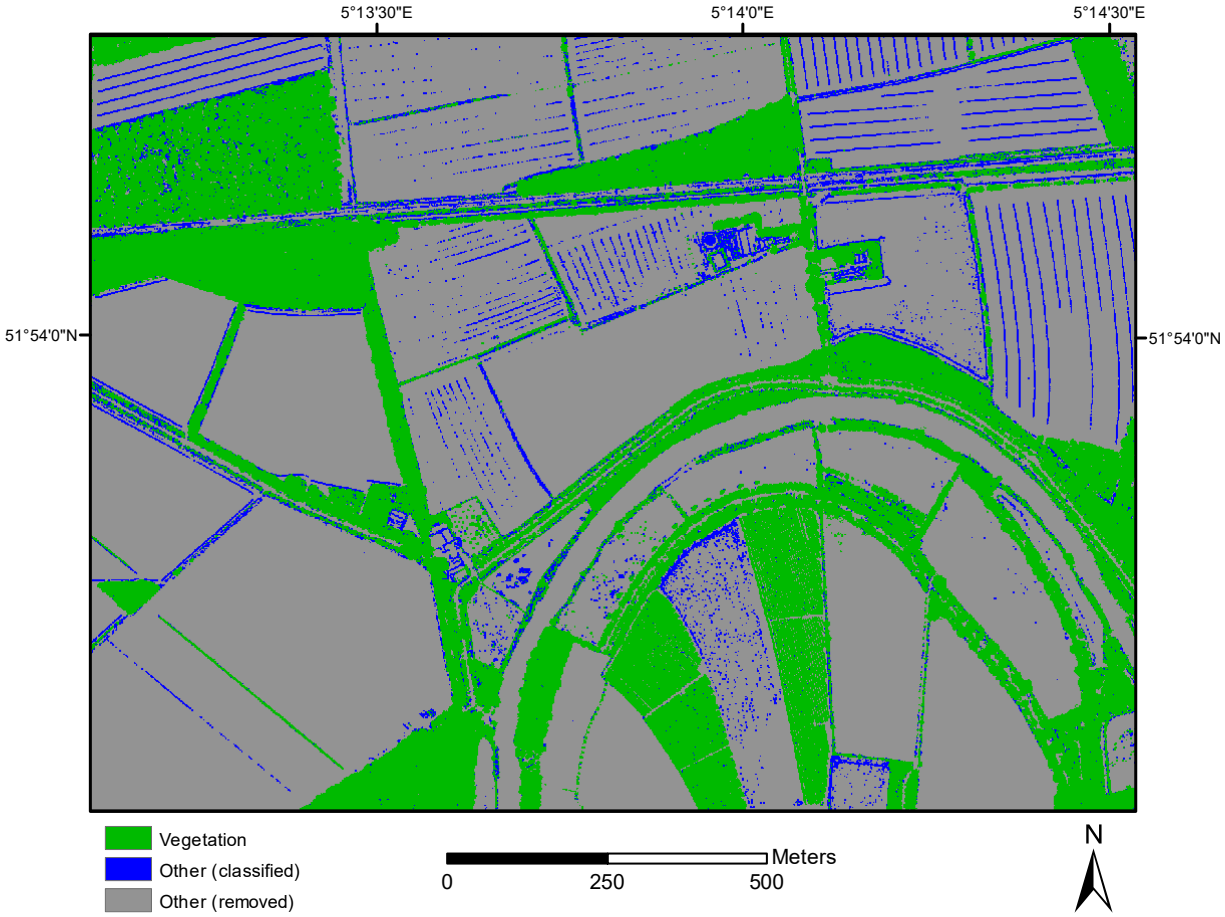

Figure 6: Results of the vegetation classification. Grey areas represent data points which were removed during the preprocessing, blue represent data points which were not removed during the preprocessing, but classified as non-vegetation (mainly building edges, ditches and railroad infrastructure). Green represents the higher than herb vegetation.

Table 3: Confusion matrix listing the automatically segmented against the manually annotated set of linear and non-linear vegetation objects in area (m^2).

		Predicted	
		Linear	Nonlinear
Actual	Linear	1159762	23438
	Nonlinear	6416	16706

classified as linear vegetation objects are shown in light green and the regions that were accurately classified as nonlinear vegetation objects are displayed in dark green. Nonlinear areas that are classified as linear are shown in red and linear regions that are classified as nonlinear are shown in orange. A confusion matrix is used to quantify the accuracy of the automated segmentation (table 3).

.. an overall accuracy of 0.87, F1 scores of 0.77, kappa of 0.70, and MCC of 0.70 respectively. Although there are some obvious errors, the majority of linear woody objects have been successfully segmented, with with users and producers accuracies of .. and .., respectively. Non-linear objects are successfully separated well, with users and producers accuracies of 0.90 and 0.93, respectively.

5. Discussion

5.1. Vegetation classification

Trimming the data based on the sphericity feature proved an efficient step to reduce the computation time needed to classify the vegetation. A substantial part of the point cloud corresponding mainly to smooth and planar areas such as bare soil, grassland and water bodies, were removed. This preprocessing step made the dataset imbalanced, but proper steps could be taken to handle the problems when classifying such datasets. When analyzing the classification statistics it is important to take this filtering step into consideration. The removed points are the ones easy to classify as not belonging to higher than herb vegetation. Consequently the remaining points that do not belong to higher than herb vegetation share many similarities with the vegetation points and are therefore harder to classify correctly. The filtered points are not part of the accuracy assessment and therefore these statistics might give a distorted picture.

Nevertheless, the majority of points were correctly classified (table 2). Most of the incorrectly classified points are dispersed and surrounded by correctly classified points. These scattered points did not have a major impact on the subsequent segmentation of linear objects.

5.2. Linear object segmentation

The comparison of the manual with automated delineation shows that linear objects were accurately extracted (figure ..), which is supported by the accuracy scores. However, the linear objects at 1 and 2 (figure ..) show profound omissions when compared to the manual classification. Closer inspection of the

rectangular objects in this area (figure ..) clarifies that a series of rectangular, but disconnected objects have been segmented instead of a single object. These objects are by themselves nonlinear and were not merged, because they are not aligned. This omission is attributed to the local pattern of the planted trees in rows of two to three across a local dike. Evidently, the objects are not always merged correctly. The merging is done based on the orientation and alignment of two adjacent and elongated objects, which works well in most cases, but not in this case. Separated single trees are also a problem, because a single tree has no meaningful orientation. Consequently a row of separated single trees will not be merged and subsequently not be classified as a linear vegetation object. The omissions at 3, 4 and 5 are the result of the downsampling of the points, since this can cause small objects to be ignored. The forest patches at 7 and 8 were automatically classified as linear elements, because of side-effects along the boundary of the study area, but would otherwise be classified correctly.

6. Concluding remarks

At present, LiDAR datasets still differ in quality, content and accessibility across and within countries. Therefore, object identification methods developed should overcome these inconsistencies. The quality of the AHN3 dataset of the Netherlands is sufficient to correctly identify linear vegetation objects with our method. In addition, multi-temporal LiDAR datasets can effectively be analyzed for change in the spatial distribution of linear vegetation objects using such a generic classifier. Initiatives to upscale the classification of linear vegetation objects, reed beds and selected forest metrics to national and European scale, based on classification of LiDAR point clouds, and using efficient cloud computing facilities are being made (Kissling et al., 2017).

The ecological value of providing such large a dataset of linear vegetation objects lies in the broad extent and fine-scale locational details, which is a powerful quality that can be used in the (3D) characterization of ecosystem structure. Existing ecosystem and biodiversity assessment projects, such as the MAES (Mapping and Assessment of Ecosystems and their Services) project (Maes et al., 2013), the SEBI (Streamlining European Biodiversity Indicators) project (Biała et al., 2012), and the high nature value farmland assessment (Paracchini et al., 2008) on a European scale and assessments of Planbureau voor de Leefomgeving (PBL) on a national level (Bouwma et al., 2014), could profit from the new details.

Acknowledgments

This work is part of the eEcoLiDAR project, eScience infrastructure for Ecological applications of LiDAR point clouds (Kissling et al., 2017), funded by Netherlands eScience Center (`https://www.esciencecenter.nl`).

References

Aguirre-Gutiérrez, J., Kissling, W. D., Carvalheiro, L. G., WallisDeVries, M. F., Franzén, M., Biesmeijer, J. C., 2016. Functional traits help to explain half-century long shifts in pollinator distributions. Scientific reports 6, doi: 10.1038/srep24451.

AHN, 2016. Inwinjaren AHN2 & AHN3. `http://www.ahn.nl/common-nlm/inwinjaren-ahn2--ahn3.html` [Online: accessed April 2017].

Aksoy, S., Akçay, H. G., Wassenaar, T., 2010. Automatic mapping of linear woody vegetation features in agricultural landscapes using very high resolution imagery. IEEE Transactions on Geoscience and Remote Sensing 48 (1), 511–522, doi: 10.1109/TGRS.2009.2027702.

Bailly, J., Lagacherie, P., Millier, C., Puech, C., Kosuth, P., 2008. Agrarian landscapes linear features detection from lidar: application to artificial drainage networks. International Journal of Remote Sensing 29 (12), 3489–3508.

Biała, K., Condé, S., Delbaere, B., Jones-Walters, L., Torre-Marín, A., 2012. Streamlining european biodiversity indicators 2020. Tech. Rep. 11/2012, European Environment Agency.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R. Q., van der Meer, F., van der Werff, H., van Coillie, F., et al., 2014. Geographic object-based image analysis–towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing 87, 180–191.

Boutin, C., Jobin, B., Bélanger, L., Baril, A., Freemark, K., 2001. Hedgerows in the farming landscapes of canada. Hedgerows of the World: their ecological functions in different landscapes, 33–42.

Bouwma, I., Sanders, M., op Akkerhuis, G. J., Onno Knol, J. V., de Wit, B., Wiertz, J., van Hinsber, A., 2014. Biodiversiteit bekeken: hoe evalueert en verkent het PBL het natuurbeleid? Tech. Rep. 924, Planbureau voor de Leefomgeving.

Bradley, A. P., 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern recognition 30 (7), 1145–1159.

Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A., 1984. Classification and regression trees. CRC press.

Burel, F., 1996. Hedgerows and their role in agricultural landscapes. Critical reviews in plant sciences 15 (2), 169–190.

Chehata, N., Guo, L., Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 38 (3), 207–2012.

Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data. Tech. Rep. 666, Department of Statistics, UC Berkeley.

Congalton, R. G., Green, K., 2008. Assessing the accuracy of remotely sensed data: principles and practices. CRC press.

Croxton, P., Hann, J., Greatorex-Davies, J., Sparks, T., 2005. Linear hotspots? the floral and butterfly diversity of green lanes. Biological conservation 121 (4), 579–584, doi: 10.1016/j.biocon.2004.06.008.

Eitel, J. U., Höfle, B., Vierling, L. A., Abellán, A., Asner, G. P., Deems, J. S., Glennie, C. L., Joerg, P. C., LeWinter, A. L., Magney, T. S., et al., 2016. Beyond 3-d: The new spectrum of lidar applications for earth and ecological sciences. Remote Sensing of Environment 186, 372–392.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Kdd. Vol. 96. pp. 226–231.

Freeman, H., Shapira, R., 1975. Determining the minimum-area encasing rectangle for an arbitrary closed curve. Communications of the ACM 18 (7), 409–413.

Gobster, P. H., Nassauer, J. I., Daniel, T. C., Fry, G., 2007. The shared landscape: what does aesthetics have to do with ecology? Landscape ecology 22 (7), 959–972, doi: 10.1007/s10980-007-9110-x.

Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. ISPRS Journal of Photogrammetry and Remote Sensing 66 (1), 56–66.

He, H., Garcia, E. A., 2009. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering 21 (9), 1263–1284, doi: 10.1109/TKDE.2008.239.

Ho, T. K., 1998. The random subspace method for constructing decision forests. IEEE transactions on pattern analysis and machine intelligence 20 (8), 832–844.

Hoppe, H., DeRose, T., Duchampt, T., McDonald, J., Stuetzle, W., 1992. Surface reconstruction from unorganized points. Computer Graphics 26, 2.

Hsu, C., Chang, C., Lin, C., 2003. A practical guide to support vector classification.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python. `http://www.scipy.org/` [Online: accessed April 2017].

Jongman, R., 2004. Landscape linkages and biodiversity in european landscapes. The new dimensions of the European landscape. Springer, Dordrecht, 179–189.

Kashani, A. G., Olsen, M. J., Parrish, C. E., Wilson, N., 2015. A review of lidar radiometric processing: From ad hoc intensity correction to rigorous radiometric calibration. Sensors 15 (11), 28099–28128.

Kissling, W. D., Seijmonsbergen, A., Foppen, R., Bouten, W., 2017. eecolidar, escience infrastructure for ecological applications of lidar point clouds: reconstructing the 3d ecosystem structure for animals at regional to continental scales. Research Ideas and Outcomes 3, e14939, doi: 10.3897/rio.3.e14939.

Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai. Vol. 14. Stanford, CA, pp. 1137–1145.

Kubat, M., Holte, R. C., Matwin, S., 1998. Machine learning for the detection of oil spills in satellite radar images. Machine learning 30 (2-3), 195–215.

Lefsky, M. A., Cohen, W. B., Parker, G. G., Harding, D. J., 2002. Lidar remote sensing for ecosystem studies: Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately

estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. AIBS Bulletin 52 (1), 19–30.
Lim, K., Treitz, P., Wulder, M., St-Onge, B., Flood, M., 2003. Lidar remote sensing of forest structure. Progress in physical geography 27 (1), 88–106.
López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences 250, 113–141.
Maes, J., Teller, A., Erhard, M., Liquete, C., Braat, L., Berry, P., Egoh, B., Puydarrieux, P., Fiorina, C., Santos, F., et al., 2013. Mapping and assessment of ecosystems and their services. Tech. Rep. EUR 27143 EN, Joint Research Center - Institute for Environment and Sustainability.
Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing 66 (6), S71–S84, doi: 10.1016/j.isprsjprs.2011.09.008.
Marquer, L., Gaillard, M.-J., Sugita, S., Poska, A., Trondman, A.-K., Mazier, F., Nielsen, A. B., Fyfe, R. M., Jönsson, A. M., Smith, B., et al., 2017. Quantifying the effects of land use and climate on holocene vegetation in europe. Quaternary Science Reviews 171, 20–37.
Matthews, B. W., 1975. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure 405 (2), 442–451.
McKinney, W., et al., 2010. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. Vol. 445. Austin, TX, pp. 51–56.
Meyer, B. C., Wolf, T., Grabaum, R., 2012. A multifunctional assessment method for compromise optimisation of linear landscape elements. Ecological Indicators 22, 53–63.
Moreira, A., Santos, M. Y., 2007. Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. In: GRAPP 2007. INSTICC Press (Institute for Systems and Technologies of Information, Control and Communication).
Nagao, M., Matsuyama, T., 2013. A structural analysis of complex aerial photographs. Springer Science & Business Media.
Paracchini, M. L., Petersen, J.-E., Hoogeveen, Y., Bamps, C., Burfield, I., van Swaay, C., 2008. High nature value farmland in europe. Tech. Rep. EUR 23480 EN, Joint Research Center - Institute for Environment and Sustainability.
Pauly, M., Gross, M., Kobbelt, L. P., 2002. Efficient simplification of point-sampled surfaces. In: Proceedings of the conference on Visualization'02. IEEE Computer Society, pp. 163–170.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.
Preparata, F. P., Shamos, M., 1985. Computational geometry: an introduction. Springer Science & Business Media.
Quackenbush, L. J., 2004. A review of techniques for extracting linear features from imagery. Photogrammetric Engineering & Remote Sensing 70 (12), 1383–1392.
Rabbani, T., Van Den Heuvel, F., Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. International archives of photogrammetry, remote sensing and spatial information sciences 36 (5), 248–253.
Scholefield, P., Morton, D., Rowland, C., Henrys, P., Howard, D., Norton, L., 2016. A model of the extent and distribution of woody linear features in rural great britain. Ecology and evolution 6 (24), 8893–8902, doi: 10.1002/ece3.2607.
Song, J.-H., Han, S.-H., Yu, K., Kim, Y.-I., 2002. Assessing the possibility of land-cover classification using lidar intensity data. International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences 34 (3/B), 259–262.
Spellerberg, I. F., Sawyer, J. W., 1999. An introduction to applied biogeography. Cambridge University Press, doi: 10.1086/393452.
Stoate, C., Boatman, N., Borralho, R., Carvalho, C. R., De Snoo, G., Eden, P., 2001. Ecological impacts of arable intensification in europe. Journal of environmental management 63 (4), 337–365.
Sun, Y., Wong, A. K., Kamel, M. S., 2009. Classification of imbalanced data: A review. International Journal of Pattern Recognition and Artificial Intelligence 23 (04), 687–719.
Tansey, K., Chambers, I., Anstee, A., Denniss, A., Lamb, A., 2009. Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas. Applied geography 29 (2), 145–157, doi: 10.1016/j.apgeog.2008.08.004.
Thornton, M. W., Atkinson, P. M., Holland, D., 2006. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. International Journal of Remote Sensing 27 (3), 473–491, doi: 10.1080/01431160500207088.
Toussaint, G. T., 1983. Solving geometric problems with the rotating calipers. In: Proc. IEEE Melecon. Vol. 83. p. A10.
Turner, M. G., 1989. Landscape ecology: the effect of pattern on process. Annual review of ecology and systematics, 171–197.
Van der Zanden, E. H., Verburg, P. H., Mücher, C. A., 2013. Modelling the spatial distribution of linear landscape elements in europe. Ecological indicators 27, 125–136.
Vannier, C., Hubert-Moy, L., 2014. Multiscale comparison of remote-sensing data for linear woody vegetation mapping. International journal of remote sensing 35 (21), 7376–7399, doi: 10.1080/01431161.2014.968683.
Vosselman, G., 2013. Point cloud segmentation for urban scene classification. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 1 (2), 257–262.
Walt, S. v. d., Colbert, S. C., Varoquaux, G., 2011. The numpy array: a structure for efficient numerical computation. Computing in Science & Engineering 13 (2), 22–30.
Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing 105, 286–304.
West, K. F., Webb, B. N., Lersch, J. R., Pothier, S., Triscari, J. M., Iverson, A. E., 2004. Context-driven automated target

456 detection in 3d data. In: Defense and Security. International Society for Optics and Photonics, pp. 133–143.
457 Yan, W. Y., Shaker, A., El-Ashmawy, N., 2015. Urban land cover classification using airborne lidar data: A review. Remote
458 Sensing of Environment 158, 295–310.