Oblique view individual tree crown delineation

Christian Kempf [a,b,*], Jiaojiao Tian [a], Franz Kurz [a], Pablo D'Angelo [a], Thomas Schneider [b], Peter Reinartz [a]

[a] *DLR, Department of Photogrammetry and Image Analysis, Oberpfaffenhofen, Germany*
[b] *TUM, Department of Ecology and Ecosystem Management, Freising, Germany*

A B S T R A C T

Individual tree crown (ITC) segmentation supports numerous applications in forest management and ecology. In the latter context, special attention is dedicated to the study of angular reflection effects, caused by the interaction of incident sunlight with a canopy. High precision airborne analysis of these effects requires multi-view sensor systems and multi-view ITC segmentation. In particular oblique view image segmentation is difficult and has been addressed by numerous template based methods. This contribution identifies persistent shortcomings in the state of the art and tackles the problem by a multi-step workflow utilizing the digital surface model (DSM), derived from multi-view stereo data. A slightly revised version of the previously published levelset-watershed segmentation of the DSM is presented as the first step. In the second step, the contour of the visible part of a candidate tree in images with known orientation is obtained by means of ray casting and concave hull calculation. The method was tested on a deciduous, mixed and coniferous plot, whose aerial images were acquired using the 3K camera system in 2018 at Kranzberg Forest, Bavaria, Germany. Accuracies were assessed at hand of human operator generated groundtruth tree tops for the DSM as well as images with zenit angles of approximately $0, 45$ and 52 degrees. The resulting F_1-scores, averaged over the plots, are $0.909/0.902/0.886/0.876$ for the DSM/near-nadir-/oblique-/maximum-oblique-images, respectively.

1. Introduction

Individual tree crown segmentation emerged in forest inventory and management (White et al., 2016). It is the foundation to collect single tree parameters (location, height, crown diameter) (Popescu et al., 2003) and to harness object based spectral analysis (Asner and Martin, 2009). Moreover, it enables to derive structural parameters, to estimate biomass (Koch, 2010) and to map carbon (Coomes et al., 2017) on the stand level and beyond using allometry (Jucker et al., 2017). The present article is focusing on the segmentation of individual tree crowns in oblique view aerial images. The need arises in forest ecology, where directional reflection properties are studied (Chen and Leblanc, 1997; Asner, 2000; Schaepman, 2007). Specifically, leveraging the approximation of the bidirectional reflectance distribution function (BRDF, (Nicodemus et al., 1977)) of a single tree requires its segmentation from several, in particular also oblique views. The principle of the proposed method can be brought forward to also differentiate sunlit and shaded sections of canopies (Martin et al., 2020).

LiDAR is already offered as operational solution in forestry applications and is able to capture the tree structure below the canopy. Nevertheless, tradeoffs between cost and quality are discussed (Jakubowski et al., 2013). Aerial optical imagery is considerably less expensive and thus to be favoured regarding feasible repetition rates of campaigns, based on a fixed budget. Frequent measurements are necessary to capture phenological behaviour, which has early been demonstrated to facilitate species discrimination in airborne, very high resolution images (Key et al., 2001) and later on using satellite imagery (Elatawneh et al., 2013). Moreover, special multicamera photogrammetric systems such as 3 K (with one nadir and two off-nadir looking cameras) can increase the performance of dense stereo reconstructions (Leitloff et al., 2014). Finally, off-nadir LiDAR measurements severely affect forest structure metrics (Liu et al., 2018).

Also oblique image acquisition is known to introduce problems, caused by projective - geometric effects, such as occlusions and scale effects, as well as (radiometric) BRDF effects (Remondino and Gerke, 2015; Gerke and Vosselman, 2016). On the contrary to the fact that the
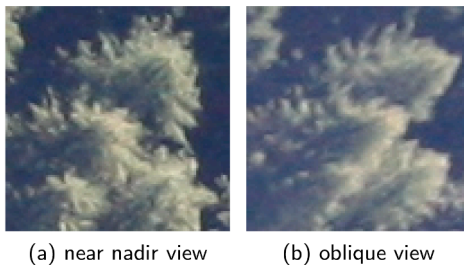

(a) near nadir view   (b) oblique view

Fig. 1. In both views, shadow and an illuminated parts of a tree can be distinguished. In the oblique view, trees become subject to occlusion.

BRDF poses a problem to e.g. orthoimage generation, its knowledge allows to draw conclusions about the properties of a surface, such as roughness and constituent materials (Chen et al., 2003). In canopy spectroscopy, vegetation parameters are inferred by inversion of canopy spectra (Ferreira et al., 2018). Closely related and also known as ray-tracing in computer graphics, forward simulations, e.g. (Ward, 1992; Gastellu-Etchegorry et al., 2017), use also BRDF-databases of materials, e.g. (Hueni et al., 2009), which are often built using ground-based goniospectrometer measurements (Fairchild et al., 1990; Roosjen et al., 2012).

Correct BRDF approximation requires accurately delineated tree crowns and knowledge of shadow areas within the boundaries. Consequently, segmentation should take into account the following effects, which are exemplified in Fig. 1:

- Occlusions, that originate from the object itself or from other objects in its vicinity, i.e. the camera's line of sight to a given object point intersects another object point that is closer to the camera. Those are even more critical, since neighbouring objects likely have similar color and texture.
- Cast shadow, that - again - originates from the object itself or from objects in the vicinity, but may occur as superposition of both cases as well. Geometric (hard) shadow points can be determined as an occlusion along the sun's line of sight towards the point in question. However, since trees exhibit scattering, and due to directional-hemispheric illumination, hard shadow detection is insufficient to detect the true tree crown boundary in the presence of shadows.

It should be noted, that the present contribution is solely dedicated to occlusion handling. Besides occlusion and shadow casting, the variability of the shape of a tree crown under different projections poses an additional difficulty: While the transformation of a shape on a plane due to a change of the camera parameters can be derived using the homography condition (Zhuo et al., 2019), this does not hold for tree contours. One could argue that a limited number of projections is generally sufficient to reconstruct a 3D object up to a limited level of detail. But this would require knowledge about the scene content and in particular the symmetry properties of the contained objects in advance.

1.1. Related work

Numerous approaches for ITC detection and segmentation have been developed in the past decades. This section is focused on fully automatic methods that explicitly address oblique views or appear suitable for adaptation.

Raytraced Templates. The first and widely adopted idea is the template matching (TM) approach of Pollock (Pollock, 1994; Pollock, 1996): A tree crown model is defined as the generalized ellipsoid of revolution (EOR). In the supervised training step, instances of the model, i.e. a set of EORs with fixed parameters, are generated. Rendering image templates

presumes initialization of the sensing geometry, scene irradiance, tree crown shape and size, reflectance and scattering properties of trees as well as background intensity. Template matching is realized using least squares, while weighting pixels by distance from the crown edge and testing lack of fit as well as significance of the regression. The optimal configuration of resulting recognition instances is found using Simulated Annealing, whose objective function avoids overlaps and prefers high correlations. (Larsen and Rudemo, 1998) extends the Pollock model in the aspect of modeling the ground around a tree by a plane, such that shadow that is casted by a tree itself, whose extent exceeds the ground projected area of the tree, is included in the template image. The precondition of a priori known number of trees for the matching procedure in this approach was attempted to overcome by template voting (Larsen, 1999).

Stochastic Models A Marked Point Process (MPP) is used in (Perrin et al., 2005): Tree objects are modeled as the marks, namely ellipses parameterized by the location, the lengths of the principal axes and its orientation. A prior term penalizes overlaps and rewards regular alignments. The latter requires an additional hard constraint on the minimum distance, to ensure stability. The likelihood term (of the image data given a configuration) is modeled as a gaussian mixture of foreground (bright pixels) and background (dark pixels). Energy minimization is performed using Reverse Jump Markov Chain Monte Carlo (RJMCMC) to sample the MPP, while it incorporates a Simulated Annealing scheme. (Descombes and Pechersky, 2006) make use of a pairwise Markov Random Field (MRF) with three labels, indicating background, vegetation and a tree centre, respectively. A pairwise MRF is a tuple of random fields which is markovian, but each one alone is not. This allows to segment images without model approximation of the prior, as compared to hidden MRFs (Pieczynski and Tebbache, 2000). The model is defined by two concentric ellipses, which represent the tree crown and its surrounding background, respectively. The prior term is the sum of a penalty for overlapping inner ellipses and a data-fitting term which distinctively evaluates the inner and outer ellipsis. As in the previously mentioned work, the likelihood is modeled as gaussian mixture, and similarly optimization is carried out using MCMC embedded into Simulated Annealing. Perrin et al. (2006) extend their previous work with concentric ellipses as well as ellipsoids with shadows, that are projected onto a plane along the direction of the sunlight. The data term is adapted by means of the Bhattacharyya distance. MPPs and MRFs are utilized in (Tolpekin et al., 2010 and Ardila et al., 2011), respectively, for superresolution mapping of trees in very high resolution (VHR) satellite imagery of urban areas.

Variational Methods. Multitemporal tree crown monitoring using coregistered red and near infrared (NIR) band images is performed in (Ardila et al., 2012). For the first image of the time series, the procedure is as follows: Initially, tree crowns are detected by fitting an elliptical Gaussian to vegetated regions in the normalized difference vegetation index (NDVI) image. The vegetated regions are found by aggregating pixels, starting from peaks in the NDVI image. The gaussian filter parameters are adapted to the area of the resulting regions. Ellipses are then simultaneously optimized using localized active contours (Lankton and Tannenbaum, 2008). (Polewski et al., 2015) reports on dead tree detection in colour infrared (CIR) imagery. Candidate regions are detected using a bayesian three component gaussian mixture model (GMM), corresponding to living vegetation, dead trees and shadows. In more detail, dead tree pixels are those, whose maximum component of the a posteori probability (found through expectation maximization) corresponds to the a priori probability of the so called "snags". Thereafter an iterative levelset segmentation with shape and intensity priors is carried out. The priors are introduced by bayesian reformulation of the energy term (Cremers, 2007), while the training shapes are represented by the eigenmodes of their signed distance functions (Tsai et al., 2003).
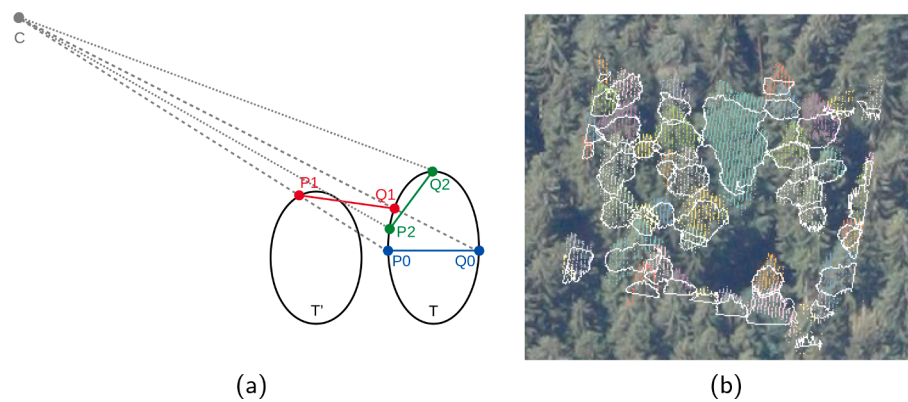
(a)   (b)

Fig. 2. (a) Simple (red) versus visible (green) boundary transformation for the 3D boundary (blue) of an ellipsoidal tree model (black). (b) An oblique image (zenit angle $\approx 40°$) with projected DSM boundaries (white) and visible projected points (coloured). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Starting points are determined as the pixels with highest dead tree propability within circles (sized according to training shapes) that cover the candidate regions.

Miscellaneous. In (Gomes and Maillard, 2014; Gomes and Maillard, 2016; Gomes et al., 2018), raytraced templates (without consideration of scattering (Whitted, 1980)) are integrated into MPPs in order to detect semi-isolated trees in satellite imagery. A pulse coupled neural network (Johnson and Padgett, 1999) gave good results in (Li et al., 2009) for trees in a sparsely vegetated savanna utilizing CIR imagery, but was unable to decompose tree clusters. (Zhao et al., 2018) compares two convolutional neural networks (UNet (Ronneberger et al., 2015), MaskRCNN (He et al., 2017)) for UAV images of a plantation.

Comparing local maxima detection, valley following, region-growing, template matching, scale-space theory and techniques based on stochastic frameworks for six different forest types (from homogeneous plantation, area with isolated tree crowns, to an extremely dense deciduous forest), (Larsen et al., 2011) concludes that none of the algorithms alone could successfully analyse all different cases. Still, region growing (Erikson, 2003) gave the best results for unmangaged, dense, mixed forests, which are in the focus of our investigations.

Summary. Prior developments do not model occlusions originating from neighbouring trees explicitly. The same holds for shadows, which are at most modeled as casted by a single tree on flat and level ground. Thus, topography of the canopies is ignored, which is insufficient for densely forested areas. Another problem is texture handling, which is only addressed in (Pollock, 1994). However, the regression analysis seeks to find the texture induced matching error merely on the basis of grey values, whereas texture clearly has a pattern with spatial variation and its saliency depends on the local illumination. All these non-modelled issues give rise to ambiguities, whose remedial measures, e. g. maximization of an objective function or template voting only have limited success. On the one hand, TM, MPP and MRF methods only detect trees and do not produce precise boundaries. On the other hand, in spite of good demonstrated performance in respective study sites, it seems tedious to devise suitable training strategies for levelsets with priors or CNNs regarding multiple views of heterogenous areas like unmanaged forests.

In the field of computer graphics, the rendering of 3D scenes, including oblique views and occlusion detection is a well studied problem. First applications in photogrammetry and remote sensing can be traced back to e.g. (Dubayah and Dozier, 1986). But also more recently these techniques have proven useful for pointcloud analysis (Alsadik et al., 2014) and object detection (Ji et al., 2019). This motivates a two step approach for oblique view delineation, which is to be presented.

1.2. Contribution

This paper presents the first part of methods to overcome the mentioned shortcomings in a geometric fashion. The underlying idea is to segment the DSM in the first step and to deduce visible contours of each segment for a given image with known orientation parameters.[1] Comparing oblique images or disparity maps of oblique stereo pairs with corresponding DSMs, tree shapes appear to be less complex in the latter.

Simple projection of segmented DSM boundaries into an oblique image is not correct, as demonstrated in Fig. 2. Fig. 2a shows a profile view of boundary point projection: Simple transformation of $P0$ (occluded) and $Q0$ (self-occluded) leads to false points in the image, namely projections of $P1$ and $Q1$. The visible range $\overline{P2Q2}$ of T is spanned by the intersections of the outermost rays through C with T, that do not yield intersections closer to C. The projection of $P1$ falls inside the projection of T' and outside the projection of the visible range of T. The projection of $Q1$ falls inside the visible projected range, leading to visible points outside the simple projected boundary. If no occlusion was present, the visible projected range would contain the simple projected range. Instances of the bespoke cases can be observed in Fig. 2b. Different scenarios can be pictured by moving the camera point C, jointly with the associated simple projection lines of $P0, Q0$ (and intersection points $P1, Q1$) as well as visible projection lines of $P2, Q2$.

To our knowledge (Kempf et al., 2019) and this work are the first contributions since 2007 that employ DSMs generated from airborne optical images (noteworthy the first-ever with a multiview camera system) for ITC segmentation of unmanaged, densely forested areas. Using multistereo 3 K-DSMs, the developed segmentation method, which was inspired by (Wang et al., 2004) gives promising results. Therefore, concerns uttered in (Straub, 2003) about degrading results due to low resolution/quality stereo DSMs should be reconsidered.

Compared to an attempt to directly segment tree crowns in oblique images, the two step approach has some important advantages: Even if only disjoint areas of a tree are visible in the oblique view, the knowledge that these are part of the same tree is preserved. Besides, calculating visible contours is computationally less expensive than most segmentation methods that are conceivable for oblique view segmentation. This should be emphasized, as BRDF construction requires sampling of the hemisphere in reasonably small azimuth and zenith angles, which in turn demands to process large amounts of the image stack.

---

[1] Strictly speaking, the image is not segmented and hence the method is termed delineation.
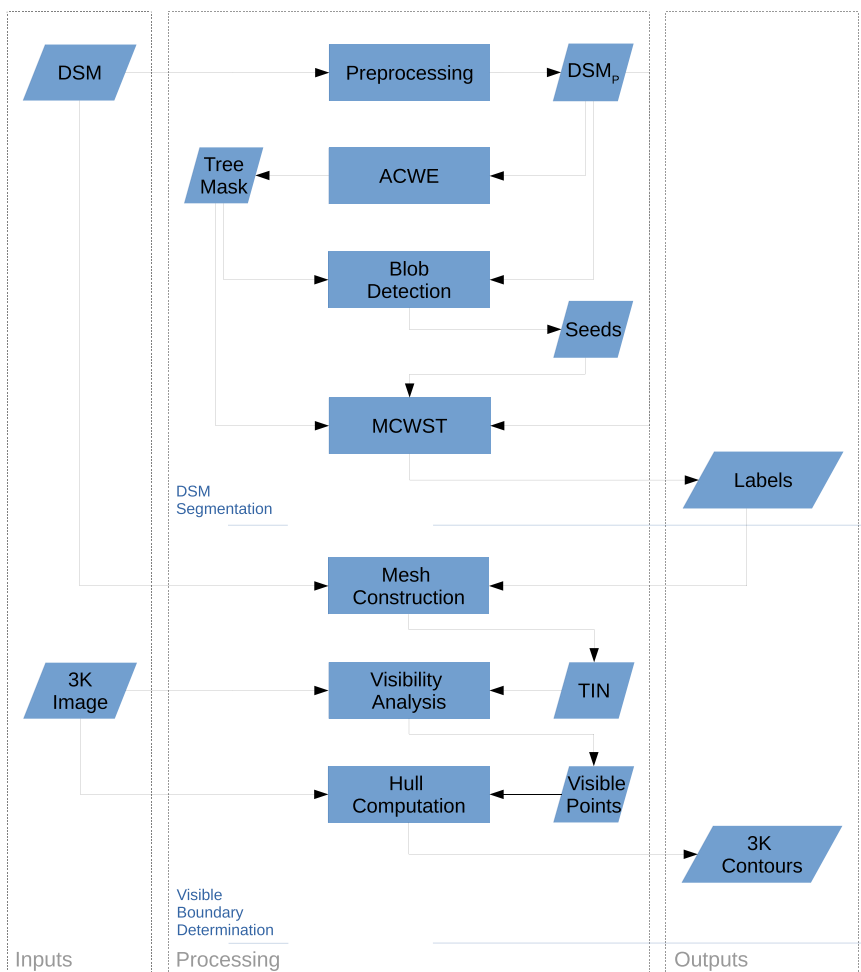
Fig. 3. Workflow for individual tree delineation in oblique images (for acronyms, see the main text of Section 1.3).

1.3. Overview

Section 2.1 describes the camera system and the processing chain for DSM generation as well as the test dataset. The workflow for oblique delineation is depicted in Fig. 3: DSM segmentation is presented in Section 2.2 and composed of Preprocessing (Section 2.2) to produce the preprocessed DSM_P, Active Contours without Edges (ACWE) (Section 2.2.1), Blob Detection (Section 2.2.2) and Marker controlled Watershed Transformation (MCWST) (Section 2.2.3). Visible Boundary Determination (Section 2.3) is subdivided into Mesh/triangulated irregular network (TIN) Construction (Section 2.3.1), Visible Point Determination (Section 2.3.2) and Hull Computation (Section 2.3.3). Sections 3.1 and 3.2 present the results and accuracy assessment of DSM segmentation and Visible Boundary Determination, respectively. The paper concludes with a discussion and evaluation of DSM Segmentation and Visible Boundary Determination in Sections 4.1 and 4.2, additionally pointing out directions for future research.

2. Material and methods

2.1. Sensor and data

2.1.1. Sensor and processing chain

The airborne optical sensor system 3K (Leitloff et al., 2014) is composed of three Canon EOS 1Ds Mark III and provides multi-view, very high resolution (VHR) images (one nadir, two off-nadir) with

several possible setups (e.g. along track, across track) in the aircraft. This allows to generate VHR DSMs. Details about the calibration and the accuracy assessment of the camera system may be found in (Kurz et al., 2007 and Kurz, 2009), respectively. Image pre-orientation is accomplished by direct georeferencing using global navigation satellite systems (GNSS) and intertial data. With overlaps found, scale invariant feature transform (SIFT) feature-points are detected, matched, filtered using random sample consesus (RANSAC) and refined using least squares. The final orientation is obtained using self calibrating bundle adjustment of the validated (SIFT-) tie points. The point cloud is calculated using semi-global matching (d'Angelo, 2016) for small-baseline images, with an overlap of at least 60%. The DSM is derived using interpolation as suggested in (Hirschmuller, 2008).

2.1.2. Test data

The test data was acquired in 2018 at Kranzberg forest, Bavaria, Germany. At the flying height of 1000 m, the 3K camera system covers an area of 2560 × 480 meters with a ground sample distance of 13 cm. Altogether, an area of approximately $4700m \times 9100m$ was captured with the 3 K camera system, which was set up in along track mode, i.e. one forward, one nadir and one backward looking camera.[2] The

---

[2] It is well known that the highest anisotropic effects occur along the sun's principal plane. Therefore, it is necessary to take flight time, camera pointing vectors and flight trajectories into account when planning the campaign.
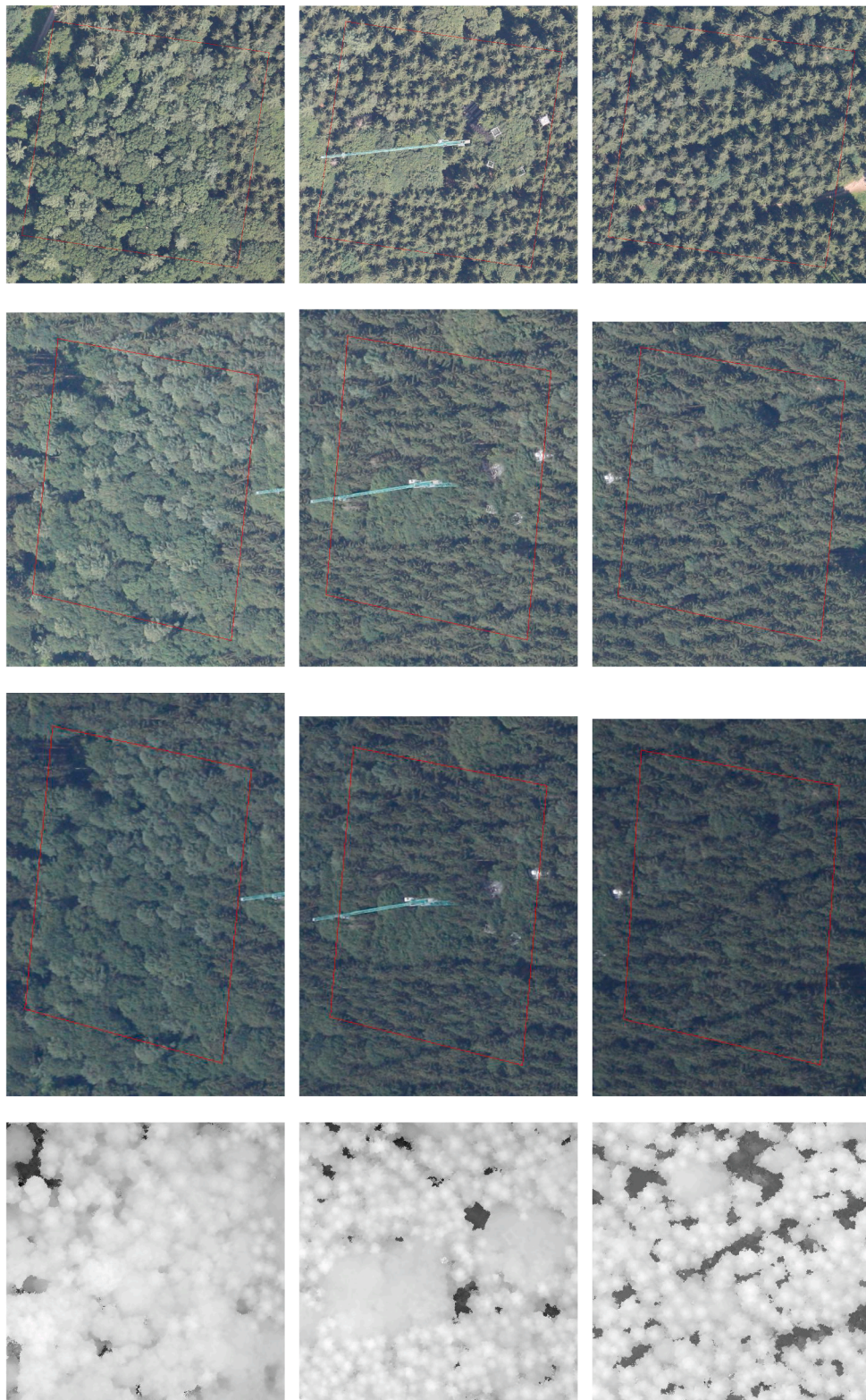
Fig. 4. From left to right: deciduous, mixed and coniferous test plots in the vicinity of the KROOF area. From top to bottom: $\approx 0°/45°/52°$ zenit angle images (with overlaid DSM-ROI) and the corresponding DSMs.

resolution of the DSM is $20cm$. Three plots (deciduous, coniferous, mixed), sized $91.2m \times 91.2m$ each, were selected in the vicinity of the KROOF project (Pretzsch et al., 2020). Referring to Fig. 4, these regions of interest (ROIs) were selected using the DSM before the corresponding image extents for 3 different zenit angles each (Table 1) were calculated by projection of the ROI corners. The experimental crane of the KROOF

site, which is located in the central mixed area has the geographic coordinates (latitude, longitude) $48°25'10.0''N$, $11°39'40.3''E$. Furthermore, three towers are situated in the test site.

The accuracy of the DSM was tested using manually marked tree tops in the image and the transformed coordinates of the DSM (Kempf et al., 2019). A comparison of nominal and actual transformed tree top

Table 1
Zenit angles of the near nadir/oblique/maximum oblique view images of each plot, given in degrees. The sign is defined negative for azimut angles in $[0, 180)$ and positive otherwise, where the azimut angle is calculated with respect to the north pointing vector.

	Deciduous	Mixed	Coniferous
image	N0125/R0110/ R0105	N0124/R0108/ R0104	N0122/R0107/ R0102
zenit angle	−1.4/44.8/52.8	−2.2/44.6/51.9	−0.9/44.3/52.5

markers showed a deviation (maximum absolute error) of less than 0.5 m. The *DSM* is given in a 2D raster representation with height $h \in \mathbb{N}$ and width $w \in \mathbb{N}$, where each cell holds a real height value and its center corresponds to a *UTM* map coordinate. Formally, the height mapping using the cell indices is denoted as $DSM : \Omega \to \mathbb{R}, \Omega = \{0, ..., w\} \times \{0, ..., h\}$, and is written $DSM[x,y] \in \mathbb{R}$. Whenever map coordinates are referred to, $DSM^{UTM}[x,y]$ is used.

2.2. DSM segmentation

Vast amounts of methods for the segmentation of optical image or LiDAR based DSMs have been published. An in-depth review would be desirable, but is out of the scope of the present article. Kempf et al. (2019) compared two variational approches to correct or circumvent that contours resulting from watershed segmentation on the inverted DSM are located in the middle of (potentially large) gaps between canopies, respectively. The approach that performed best combines Active Contours without Edges (ACWE) and Marker controlled Watershed Transformation (MCWST), henceforth abbreviated ACWE-MCWST, is revisited in the following.

The underlying idea was to improve (Wang et al., 2004) in using a non-local instead of a local method to create an approximate object mask prior to marker detection and watershed transform. Single steps of the algorithm (illustrated in Fig. 5) are enumerated below, utilizing the following definitions: The masking operation on raster arrays using an equally sized boolean matrix $B : \Omega \to \{0, 1\}$ is notated as $DSM|_B = [DSM[x,y] : B[x,y] = 1]$. Elementwise evaluation of a matrix on a number field $A \in \mathbb{K}^{w \times h}$ with respect to a relational operator $\circ_{rel} : \mathbb{K} \times \mathbb{K} \to \{0, 1\}$ is written as $A \circ_{rel} k \in \{0, 1\}^{w \times h}, (k \in \mathbb{K})$.

1. Global thresholding of the *DSM*, followed by morphological closing yields $DSM_{thr} \in \{0, 1\}^{w \times h}$.
2. Local histogram equalization (LHE) of $DSM|_{DSM_{thr}=1}$ produces DSM_{lhe}.
3. Gaussian smoothing of DSM_{lhe} gives DSM_{gs}.
4. ACWE segmentation of DSM_{gs} yields the levelset $\phi^*_{DSM} \in \{0, 1\}^{w \times h}$.
5. Blob detection on $DSM_{gs}|_{\phi^*_{DSM}=1}$ results in the set $M_{Blobs} = \{(x_i, y_i, radius_i) : i \in \mathbb{N}\}$.
6. MCWST on $DSM_{gs}|_{\phi^*_{DSM}=1}$, using the markers $M = \{(x, y) : (x, y, radius) \in M_{Blobs}\}$.

The final result is a label array, denoted *LAB*, consisting of foreground labels of MCWST and the background label. The boundaries of the foreground segments are denoted Γ_{LAB}.

A canopy height model (CHM) is actually required for the first step. The test area is relatively flat and the ground height is found manually in a preliminary examination of the data. The global threshold is set according to the minimum crown base height (e.g. 10 meters). Morphological closing with a circular structuring element with radius of 5 pixels ($\approx 1m$) is used to eliminate trees that are considered too small or noise. In the resulting foreground area $DSM|_{DSM_{thr}=1}$, LHE (Gonzalez and Woods, 2001) is carried out, with the mask size set such that it is larger than the expected minimum gap between two trees (here: 20 pixels $\approx 4m$). This is followed by one pass of gaussian smoothing with $\sigma = 1$ of DSM_{lhe} and results in DSM_{gs}. Preprocessing the DSM as just described enables the application of ACWE to DSM_{gs}, globally. Without LHE, the DSM violates the assumptions for the piecewise constant image model too heavily.

2.2.1. Active contours without edges

Let $C(q) : [0, 1] \to \mathbb{R}^2$ be a parametric curve, represented (cf. (Osher and Sethian, 1988)) as zero-levelset $C = \{(x, y) : \phi(x, y) = 0\}$ and let $D : \Omega \to \mathbb{R}^+$ the preprocessed DSM. Chan and Vese (2001) solve the piecewise constant case of the Mumford-Shah functional, which approximates D as $\widetilde{D} : \Omega \to \{v1, v2\}, (v_1, v_2 \in \mathbb{R}^2)$ with edges C. The associated energy functional is additively composed of a fitting and a regularization term. The fitting term is

$$\int_{(x,y) \in \Omega : \phi(x,y)<0} |D(x,y) - v_1|^2 dxdy + \int_{(x,y) \in \Omega : \phi(x,y)>0} |D(x,y) - v_2|^2 dxdy, \quad (1)$$

where $v_1 = average(\{D(x, y) : (x, y) \in \Omega \wedge \phi(x, y) < 0\})$, $v_2 = average(\{D(x, y) : (x, y) \in \Omega \wedge \phi(x, y) < 0\})$. The regularisation term is $\mu|C|$, where $|C|$ is the length of the boundary and $\mu \in \mathbb{R}$. The goal is to find v_1, v_2, C that minimize the sum of fitting and regularization term. The levelset representation of C assures topology adaptivity during the contour evolution. Variational formulation requires the heaviside function

$$H(\phi) = \begin{cases} 1, & \phi \geqslant 0 \\ 0, & \phi < 0, \end{cases} \quad (2)$$

such that one can rewrite $|C| = \int_\Omega |\nabla H(\phi)|$, where ∇ is the gradient operator. Here, the initial levelset ϕ^0_D is allocated as a chessboard pattern (Getreuer, 2012). The obtained final levelset ϕ^*_D serves as object mask in the forthcoming steps.

2.2.2. Blob detection

In general, a blob denotes a bright feature on dark background or vice versa. One possibility to detect such features with approximately circular shape is the Laplacian of a Gaussian (LoG). Let $g(x, y, \sigma) = \frac{1}{2\pi\sigma} e^{-\frac{x^2+y^2}{2\sigma}}$ denote a gaussian kernel. The convolution of a raster array D (the preprocessed DSM) with g at different scales σ creates a scale space representation (Lindeberg, 1994)

$$L(x, y, \sigma) = \begin{cases} D(x, y) & , \sigma = 0 \\ g(x, y, \sigma) * D(x, y)) & , \sigma > 0, \end{cases} \quad (3)$$
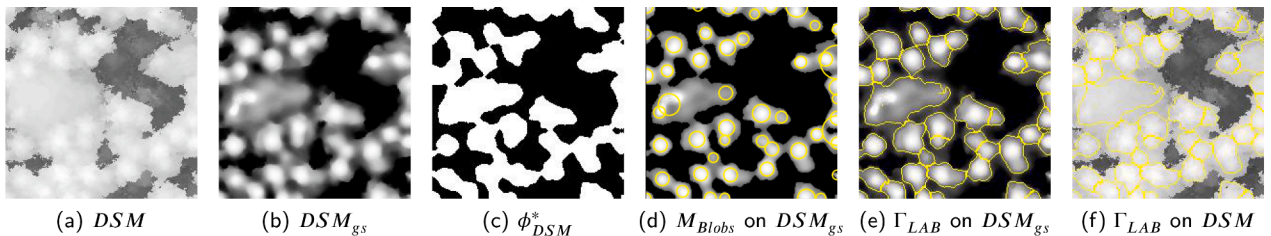

| (a) DSM | (b) DSM_{gs} | (c) ϕ^*_{DSM} | (d) M_{Blobs} on DSM_{gs} | (e) Γ_{LAB} on DSM_{gs} | (f) Γ_{LAB} on DSM |

Fig. 5. Input (a), intermediate results (b-d) and final output (e-f) of ACWE-MCWST.

(a) Γ_{LAB} on DSM (b) IMG (c) TIN (d) VIS_P on IMG (e) Γ_{VIS_P} on IMG
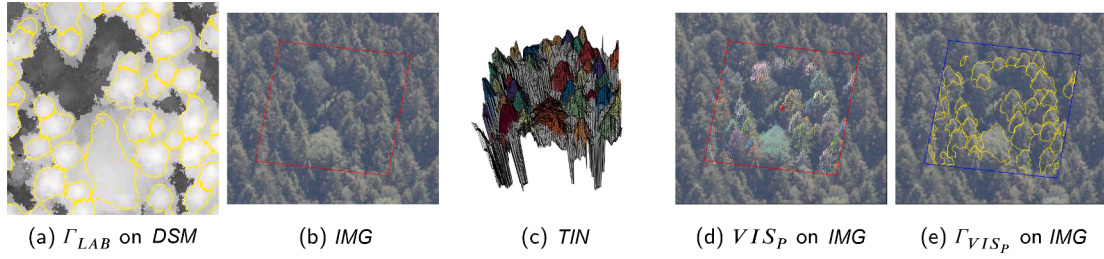
Fig. 6. Inputs (a-b), intermediate results (c-d) and final output (e) of visible boundary determination.

where $*$ denotes the convolution operator. The Laplace operator is defined as the trace of the principal axis representation H' of the Hessian matrix:

$$\nabla^2 = trace(H') = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{4}$$

The scale normalized Laplacian of a Gaussian is defined as

$$\nabla^2_{norm}g = \sigma\nabla^2 g. \tag{5}$$

Using $\nabla^2_{norm}g$, a discrete 3D-laplacian scale space can be built up. A point (x, y, σ) in this scale space is identified as a blob if it is the argument of the maximum normalized filter response among its 26 neighbours. The scale range is chosen such that it matches the expected sizes of the tree crowns and the union of blobs is referred to as M_{Blobs}. Based on the scale space maxima selection, there can still exist blobs that overlap. Brandtberg and Walter (1998) removes smaller blobs when an overlap of a certain percentage is exceeded. Since this can lead to omissions when crowns indeed overlap and it is also not evident which blobs are spurious, all blobs are retained as seeds for the subsequent MCWST in the present approach. Critically, this implies a bias towards over-segmentation. Ambiguities can be resolved using a MRF model on the detected blobs, similar to (Zhang and Sohn, 2010), for example.

2.2.3. Marker controlled watershed transformation

Visually speaking, watershed transformation (WST) mimics a flooding simulation that starts from the minima of all catchment basins in a heightmap. A watershed line is created, where the water of neighbouring segments meets. For unseeded WST, this often leads to an oversegmentation due to spurious local minimas, created by noise. Hence, seeded WST (MCWST) uses a restricted set of minima to start the flooding process. The previously introduced LoG-Blob feature detector is one possible way to find the non-spurious minima. The exact definition of watershed lines, namely the skeleton of geodesic influence zones, can be found in (Beucher and Lantuéjoul, 1979). MCWST is applied on $D_{gs}^{-1}|_{\phi^*_D=1}$, with $D^{-1} = [max(D) - D[x, y] : (x, y) \in \Omega]$, using the marker set $M = \{(x, y) : (x, y, radius) \in M_{Blobs}\}$ and yields a label array $LAB : \Omega \rightarrow Labels \cup \{0\}$, where $Labels = \{1, ..., |M|\}$ are the foreground labels (corresponding to the tree crowns) and the zero labels the background pixels (x, y) with $\phi^*_{DSM}(x, y) = 0$. The corresponding contours Γ_{LAB} are the set of boundary points of foreground segments, i.e. for $l \in Labels$, $\Gamma_{LAB}(l) = \partial LAB|_{LAB=l}$, $\Gamma_{LAB} = \cup_{l \in Labels}\Gamma_{LAB}(l)$.

2.3. Visible boundary determination

At this stage of processing, given an oriented image IMG, and the label array LAB of the segmented DSM, the visible boundary for each foreground label $l \in Labels$ in a simulated view of the DSM according to the orientation of IMG is to be deduced. Let $(x, y)^T$ be a pixel coordinate of IMG and let $(X, Y, Z)^T \in UTM$ be its corresponding point of DSM^{UTM}.

The transformation.[3] between image coordinates and world coordinates is described by the homogenous projection matrix $P \in \mathbb{R}^{4x4}$. P is composed of the camera matrix $K \in \mathbb{R}^{3x3}$ (holding the intrinsic parameters), a rotation matrix $R \in \mathbb{R}^{3x3}$ and a translation to the camera center C in world coordinates (constituting the extrinsic parameters) (Hartley and Zisserman, 2004):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \underbrace{K \begin{bmatrix} R & -RC \\ 0 & 1 \end{bmatrix}}_{P} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{6}$$

In the first place, it is required to determine the visible surface points. A lower bound for the problem of $O(n^2)$ was already proven in (McKenna, 1987). One alternative is the well known z-Buffer algorithm (Foley et al., 1996), which is based on investigation of depth values in the parallel projection of the DSM, viewed according to the image orientation. It suffers the drawbacks that depth resolution is limited to machine precision and rasterization of objects can lead to aliasing. Therefore ray-object intersection (ray casting) tests are preferred, which presume a surface representation, such as meshes. The visible boundary is defined as the concave hull of visible points. Its computation consists of the following steps and is illustrated in Fig. 6:

1. Convert the DSM^{UTM} to a triangle mesh (TIN) and assign the cell labels of DSM to the associated vertices of TIN. Retrieving the label of a vertex v is noted as $label(v)$.
2. Perform ray casting according to the orientation of IMG for each vertex of the TIN to obtain the set of visible points VIS. Let $VIS(l) = \{v : v \in VIS \wedge label(v) = l\}$ for $l \in Labels$.
3. For each $l \in Labels$, calculate the concave hull of $VIS_P(l) = \{Pv : v \in VIS(l)\}$, abbreviated $\Gamma_{VIS_P}(l)$. The set of visible boundaries is $\Gamma_{VIS_P} = \cup_{l \in Labels}\Gamma_{VIS_P}(l)$.

2.3.1. Triangle mesh generation

Let $S \subseteq \mathbb{R}^2$ be the set of points, henceforth called sites, which is obtained by dropping the height coordinate of DSM^{UTM} points. The Voronoi Region of a site $s \in S$, V_s, is defined as the set of points $x \in \mathbb{R}^2$, whose distance to s is not greater than to any other site $s' \in S$, formally:

$$V_s = \left\{ x \in \mathbb{R}^2 : s \in S \wedge \forall s' \in S(s' \neq s \Rightarrow (|x - s| \leqslant |x - s'|)) \right\}. \tag{7}$$

Points in V_s are intersections of a finite number of closed half planes and form a convex polygon. Two distinct Voronoi regions can only intersect along their boundaries. Three adjacent Voronoi regions intersect in a point. If no four sites lie on the same circle, S is said to be in general position. Moreover the union of Voronoi regions covers $\mathbb{R}^2$. The Voronoi Diagram of S is the set of Voronoi Regions for each of the sites, and was firstly described in (Voronoi, 1908).

---

[3] Here, the dependence of the exterior camera orientation on the GPS position and IMU parameters (Kurz et al., 2007) can be ignored.
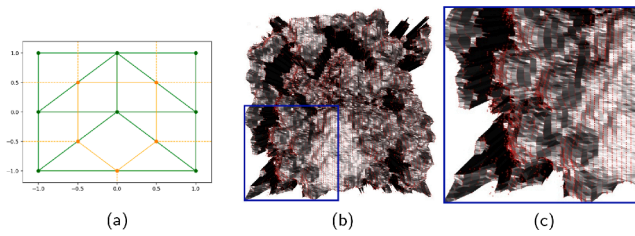
Fig. 7. (a): An example that demonstrates the construction. Input points and Delaunay edges are shown in green. The Voronoi Diagram is shown in orange (b): The Delaunay Mesh for the DSM patch. (c): A magnified region of the mesh. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Under the assumption of general position, the Delaunay Triangulation is the dual graph of the Voronoi Diagram. So given the sites *S*, the Delaunay Triangulation (Delaunay, 1934) of *S* is a graph with vertex set *S*. Its edge set is obtained from the Voronoi Diagram of *S* by creating one edge between each two sites that have adjacent Voronoi regions. The created graph is maximal planar, i.e. no additional edge can be added between any two points without crossing an existing edge. The computation for *n* points in $\mathbb{R}^2$ can be performed in $O(n\log n)$. More details on properties of triangulations and efficient algorithms may be found (Guibas et al., 1992 and de Berg et al., 2008). Fig. 7a demonstrates the construction of the Delaunay triangulation from the Voronoi diagram on a grid and Fig. 7b shows the resulting *TIN*, i.e. the Delaunay Graph with reintroduced height coordinates, for the DSM in Fig. 6a.

2.3.2. Visible point determination

A point $(X, Y, Z)^T$ of DSM^{UTM} is visible according to the orientation of the image *IMG* in the pinhole model, with camera center *C*, if the line

$$Ct + \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}(1-t), \qquad t \in [0,1], \tag{8}$$

between both points does not intersect any triangle or any other point of the previously constructed mesh. The union of visible points is denoted as the set *VIS*.

In order to solve problems like object intersection and visibility efficiently, several approaches to build a hierarchical structures were developed (Ericson, 2004). These data structures subdivide the point space to accelerate queries. One distinguishes between spatial decompositions (using e.g. binary space partitioning (BSP) trees, K-d trees or Octrees) and bounding volume hierarchies (e.g. trees of spheres, axis aligned bounding boxes (AABB), oriented bounding boxes (OBB), k - discretely oriented polytopes (k-DOP) and swept sphere volumes (SSV)). The latter are more focused on objects than the former. The enumeration of bounding volumes just mentioned is according to increasing complexity of their construction, which goes hand in hand with their fit to the object. Hence, the OBB is a well balanced, yet efficient bounding volume. The related bounding volume hierarchy is the OBB-tree and was developed in (Gottschalk et al., 1996).

An OBB is defined by the principal axes of a set of vertices, which are given by the Eigenvectors of the covariance matrix. Let *A* and *B* be two OBBs and let *T* be the vector that connects their centers. *A* and *B* are separated if there exists an axis *L*, such that the sum of their projected radii r_A and r_B is less than the distance between their projected centers, i. e. if

$$|T \cdot L| > r_A + r_B. \tag{9}$$

At most 15 axes must be tested to determine the OBB overlap status, namely those that correspond to the three faces of A, the three faces of B, and nine pairwise combinations of edges facing each other (Gottschalk et al., 1996). So, A and B are disjoint if and only if there exists one separating axis for which the axis test fails.

The OBB-tree is built up in a top-down manner: An OBB is firstly fitted to the complete TIN before nested OBBs are used to subdivide the enclosed space and thereby define a hierarchy. This principle is continued recursively, as illustrated in Fig. 8. Splitting inside an OBB
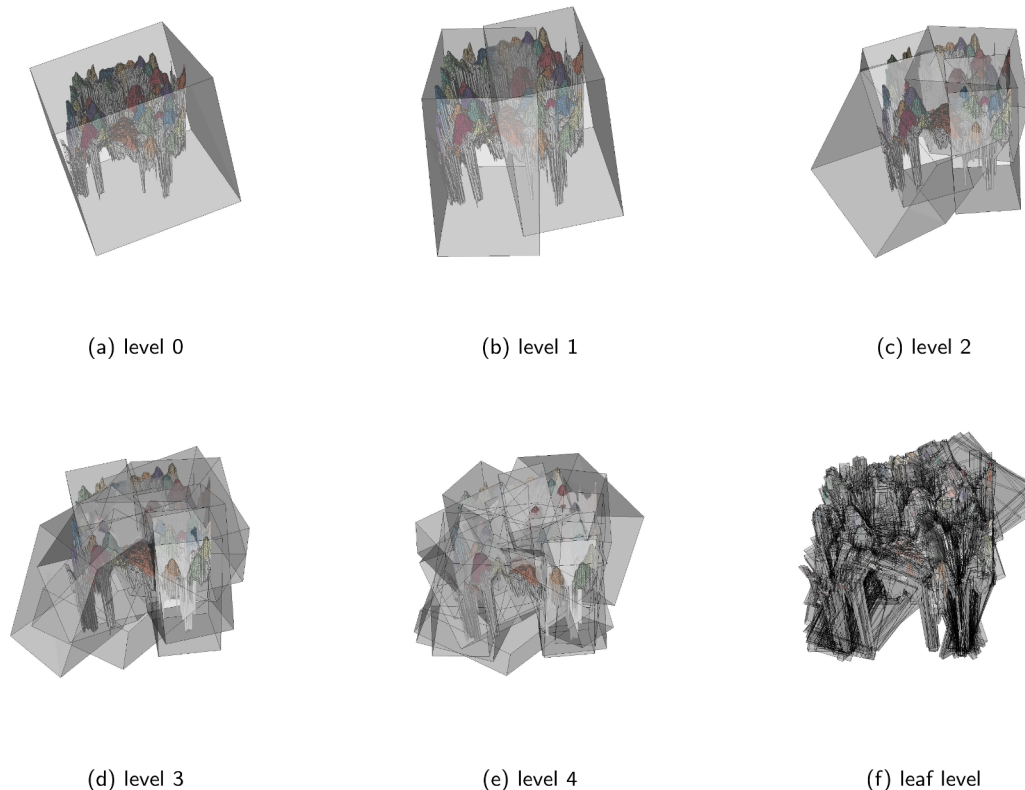

(a) level 0

(b) level 1

(c) level 2

(d) level 3

(e) level 4

(f) leaf level

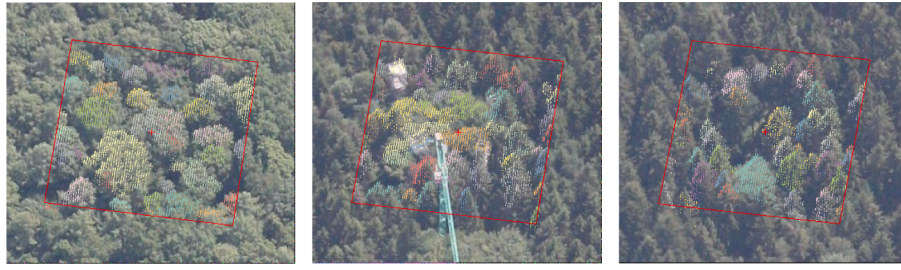**Fig. 8.** Recursive subdivision for building the OBB-tree.

Fig. 9. Visible points (coloured per tree) at zenit angle $\approx 40°$, projected onto associated images, for a subarea of the deciduous, mixed and coniferous plot (from left to right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

happens across the plane which is perpendicular to its longest axis and located at the mean of its vertices. Primitives on the splitting plane are assigned towards the side of their centroid. If a subdivision along this axis is not possible, then the second longest is chosen. If no split can be established in that manner, the set of primitives is declared indivisible. A balanced tree is created using this procedure, if the partition coordinate is set to the median center point.

The calculation of the covariance matrix and the eigenvectors from all points can result in an OBB with bad fit. Several attempts have been made to sort out the problem, e.g. using the convex hull after uniform sampling (Gottschalk et al., 1996), or a continuous formulation of covariance using triangle-wise covariance calculation with incremental matrix construction (Gottschalk, 2000), which will be considered in future work. The projected and labeled visible points for a subarea of each test plot in test images with approximately 40 degrees zenit angle are displayed in Fig. 9.

2.3.3. Hull Computation

Let $l \in Labels$ be a label and let $S \subseteq \mathbb{R}^2$ be the sites, that are obtained from $VIS(l)$ by projection,

$$S = VIS_P(l) = \left\{ P(X,Y,Z,1)^T : (X,Y,Z)^T \in VIS(l) \right\}. \tag{10}$$

The determination of the shape of a point set essentially depends on the perception which points belong to the boundary, or in other words, on the point density characterizing the interior of the recognized objects. The convex hull is independent of the point density. Mathematical preliminaries of convex sets can be found concisely in (Baerentzen et al., 2012). In particular the following definition is essential in the present context: For a set of points S, the convex hull is the intersection of half spaces containing S, formally:

$$CH(S) = \bigcap_{H \supseteq S \wedge H \text{ is a half space}} H. \tag{11}$$

It evoked investigations to replace the hyperplane with other structural elements. One of the first approaches is known as the Jarvis construction, which studied the effect of finite hyperplanes on concavities. Edelsbrunner instead introduced the use of disks, which is the basic idea of the α-shape (Edelsbrunner et al., 1983) and explained more precisely below. The concept was later on extended for weighted points in (Edelsbrunner, 1992) and 3D (Edelsbrunner and Mücke, 1994). The proven worst case times for 2D and 3D are $O(n\log n)$ and $O(n^2)$, respectively.

Different approaches were published for concave hulls: (Moreira and Santos, 2007) is using k-nearest neighbours in order to disassemble cluttered point sets, composed of multiple objects with slightly inhomogeneous point distribution. Asaeedi et al. (2013) aims for a single object by imposing a bound on the interior angles of the bounding polygon. In the present study, the point density problem can be avoided: Triangles can be split after Delaunay triangulation by considering a histogram of triangle areas. Thus, a homogeneous point distribution can be obtained, which is maintained throughout the visibility analysis step. α-shapes have been widely utilized, especially in the study of

biochemical cell structures. Some examples for applications within remote sensing and forestry are (Escalante, 2012; Ning et al., 2016 and Trochta et al., 2017; Yan et al., 2019), respectively. Now, necessary definitions in (Edelsbrunner et al., 1983) are recapitulated.[4]

Let $x \in \mathbb{R}^2$. The open disk of points, which are closer than distance r from x is denoted $D(x,r)$. Its closed complement is denoted $\overline{D()}$ and is written as

$$\overline{D(x,r)} = \left\{ x^{'} \in \mathbb{R}^2 : \left\| x^{'} - x \right\| \geqslant r \right\}. \tag{12}$$

Let $\alpha \in \mathbb{R}$. If $\alpha < 0$, the intersection of all closed complements of disks with radii $-1/\alpha$, that contain all points of S is called the α-hull of S, formally:

$$\alpha - hull(S) = \bigcap_{x \in \mathbb{R}^2} \left\{ \overline{D(x, -1/\alpha)} : \overline{D(x, -1/\alpha)} \supseteq S \right\}, (\alpha < 0). \tag{13}$$

A point of S is called α-extreme, for some $\alpha \in \mathbb{R}$, if it is on the α-hull. Two extreme points that are adjacent on the boundary of the same open disk are called α-neighbours. The α-shape is the straight line graph, whose vertices are the α-extreme points and whose edges connect the α-neighbours. Let $\alpha \neq 0$ and let F be a face of the α-shape of S. Furthermore, let $e = (p,q)$ be a boundary edge of F. For $\alpha < 0$, e is called positive edge of F, if there exists an open disk with radius $-1/\alpha$, and with p and q on its boundary, such that it (i) contains at least one point of S and (ii) has its center on the same side of e as F. Otherwise e is called negative. F is called interior (exterior), if one of its boundary edges is a positive (negative) edge of F.

Three facts are proven for vertices and edges, which harness the construction from the Delaunay triangulation: (1) The α shape for $\alpha < 0$ is a subgraph of the closest point Delaunay triangulation.[5] (2) The range of α values, for which a site of S is α-extreme is bounded. (3) The range of α values, for which a Delaunay edge is also an edge of the α-shape is bounded. Another fact is stated for faces, which ultimately describes the construction of the α-shape from the Delaunay Triangulation: (4) For $\alpha < 0$, the set of interior faces of the α-shape equals the union of triangles in the closest point Delaunay triangulation, whose circumcircle radius is not greater than $-1/\alpha$.

With respect to the outline at the beginning of Section 2.3, $\Gamma_{VIS_P}(l) = \Gamma_S = \alpha - shape(S)$.

Fig. 10 demonstrates the principle for the visible, projected points of an example segment: For $\alpha = 0$, the α-shape is the convex hull. As α is decreased, the number of triangles that are removed increases until, for a certain $\alpha > -\infty$, the α-shape becomes the set of isolated points S.

---

[4] This is restricted to the case $\alpha < 0$, which results in concave hulls. In publications that succeed the original one, the convex hull case for $\alpha > 0$ is often ignored and the definitions are adapted to handle only the concave case. Concretely, as a consequence, $\alpha > 0$ is used to define concave hulls. Here, the terminology of the original publication is adpoted.

[5] This is the variant introduced in Section 2.3.1, where the Voronoi Diagram could likewise be called closest point Voronoi Diagram.
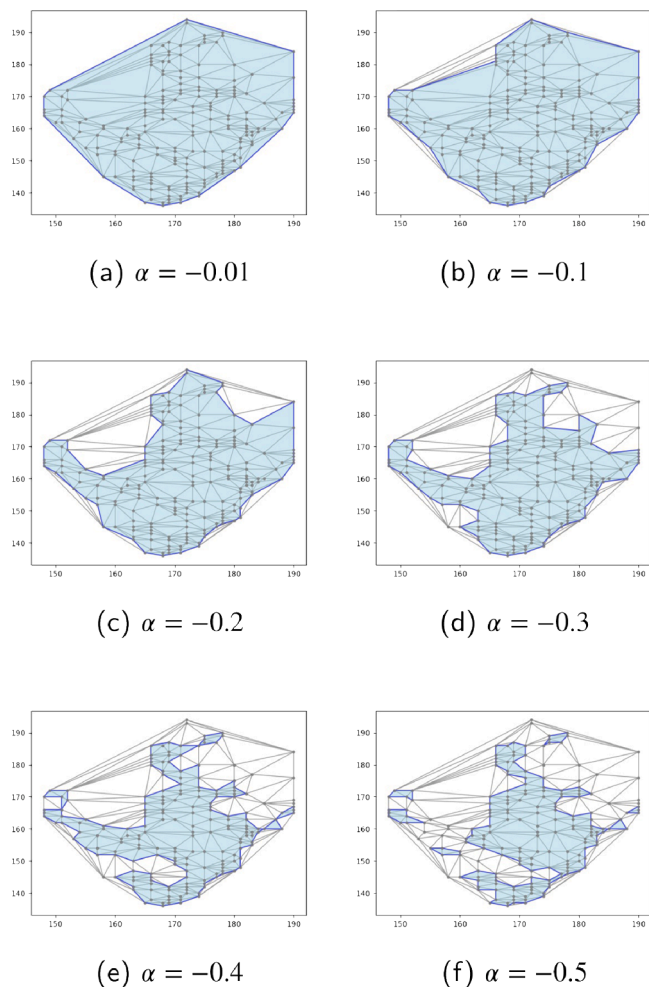
(a) $\alpha = -0.01$ (b) $\alpha = -0.1$

(c) $\alpha = -0.2$ (d) $\alpha = -0.3$

(e) $\alpha = -0.4$ (f) $\alpha = -0.5$

Fig. 10. The Delaunay triangulation is shown in gray. The alpha shape for several α-values is shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Experiment and results

3.1. DSM segmentation

ACWE-MCWST was applied to each of the test plots, that were presented in Section 2.1.2, Fig. 4. The best results were achieved for the coniferous plot, followed by the deciduous and the mixed plot (Fig. 11). One cause for errors in the mixed plot is the presence of the towers in the experimental site, which could not be excluded as done for the crane during the DSM generation. One possibility to overcome this would be to manually mask those objects. For operational applications, it seems more appropriate to exclude those object via preliminary analysis of coregistered NIR or hyperspectral images. In more general, coniferous and deciduous trees can be well segmented, but there are kinds of deciduous tree clusters where separation is not possible for the human eye and also not for the algorithm. Comparing the top and bottom row of Fig. 11, it is apparent, that segments are slightly contracted, which arises out of local histogram equalization. Albeit, it was found in (Kempf et al., 2019), that this improves the performance of the LoG-Blob detection and enables to disassemble deciduous tree clusters to a certain extent.

Table 2 summarizes the detection accuracy of the method. The groundtruths consist of tree crown tops/centers in pixel coordinates (CC), which were manually created by inspection of the surface models and corresponding orthorectified images. If a CC is contained in no segment, this is defined as omission (also false negative, FN). Otherwise, if no, one, or $n > 1$ CC(s) is (are) contained in a segment, this is defined

as a commission (also false positive, FP), detection (also true positive, TP) or combined 1-detection - (n-1)-omission, respectively.[6] Therefore, the number of CCs is $\#CC = TP + FN$ and the number of segments is $\#SG = TP + FP$. Due to pixel representation, the contingency table for the bespoke values can be built in $\#CC + \#SG$ steps.

Based thereupon, the following metrics, cf. e.g. (Fawcett, 2006) or (Tharwat, 2018), are calculated[7]:

$$
\begin{aligned}
Precision, PC &= TP/(TP + FP), \\
Recall, RC &= TP/(TP + FN), \\
F_1 - Score, F1 &= \frac{2}{1/PC + 1/RC} \\
&= 2TP/(2TP + FP + FN).
\end{aligned}
\tag{14}
$$

3.2. Visible boundary determination

The full set of intermediate results, i.e. the visible points projected onto the corresponding images is shown in Appendix B, Fig. 15. Fig. 13 shows the final results of the described method for each of the test images. The oblique visible ROI, shown in blue, is calculated from the ROI of the corresponding DSM and the visible boundaries (see Appendix A for details). The groundtruth points, shown in green and cyan, are obtained by transformation of the DSM groundtruth coordinates, translated by $-0.5m$ in the z-coordinate, into the image according to Eq. (6). No visual-manual adjustment of the groundtruth was performed. Hence, three factors influence the detection results:

- The contingencies and calculated metrics in Table 3 reflect the height errors of the DSM. Thus, a groundtruth crown center (CC) can be slightly outside its segment, despite it was detected during ACWE-MCWST in the DSM. Two cases can be distinguished: The CC can fall into a neighbouring segment and hence is counted as one omission and one commission error. In this manner, systematic height errors and the topology of segments can lead to a chain effect where errors compensate. A CC that falls outside of its segment into the background equally counts as one omission and one commission error. Projecting groundtruth points with lower z-coordinate to solve these issues is not always correct, e.g. for flat or occluded objects. Examples for the former can be seen at the right border of the deciduous plot.
- A CC might be hidden. There are again two cases (assuming that the involved CCs are within the boundary of the occluding segment): When the complete segment is hidden, this leads to an omission error. An additional commission error occurs, when parts of the occluded segment are visible. The latter happened in the upper right corner of the mixed plot.
- A close look at Fig. 15 reveals that the point density is decreasing for oblique views at surface areas with steep slopes, as present for coniferous trees. Too large α values then lead to removal of triangles, disintegrated components or even vanishing α shapes without edges. The choice of α in the case of heterogenous point density needs to compromise between exact hulls (with stronger concavities) and omissions. This problem can be solved by mesh refinement after the Delaunay triangulation in step one of the visible boundary determination. Here, it was scrutinized that no segment disappeared.

Tables 3,4 list the results of accuracy assessment of visible boundary determination and the differences to the DSM segmentation for comparison, respectively. One groundtruth point of the mixed plot, (i, j, z) = (420.24, 308.25, 529.18164), was discarded because its transformation

---

[6] The only true negative, TN, is the background label, which is not considered a segment.

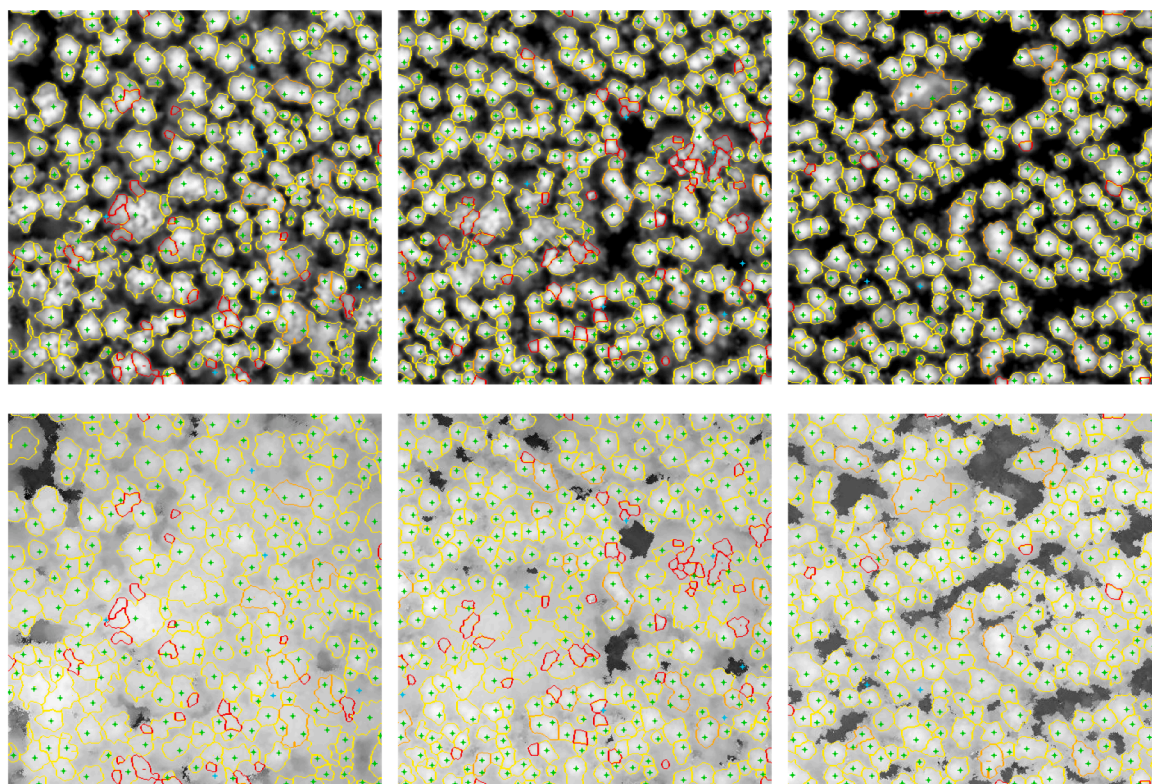[7] The F_1-score is also called F-measure.

Fig. 11. Segmented boundaries (displayed in yellow/orange/red for detection/omission/commission) and groundtruth tree tops (displayed as green/cyan markers for detection/omission) are superimposed on DSM_{gs} (top row) and DSM (bottom row) for the deciduous, mixed and coniferous plot (from left to right).

Table 2
Detection accuracy of ACWE-MCWST for the test plots.

	Deciduous	Mixed	Coniferous
total trees	205	271	224
total segments	222	298	220
detections	193	248	211
omissions	12	23	13
commissions	29	50	9
precision	0.869	0.832	0.959
recall	0.941	0.915	0.942
F_1-score	0.904	0.872	0.950

Table 3
Detection accuracy of visible boundary determination based on ACWE-MCWST segmentation of the DSM for the test plots in images with $\approx 0°/45°/52°$ zenit angle.

	Deciduous	Mixed	Coniferous
image	N0125/R0110/R0105	N0124/R0108/R0104	N0122/R0107/R0102
total trees	205/205/205	271/270/270	224/224/224
total segments	222/222/221	298/295/295	220/220/220
detections	190/188/185	247/239/236	211/207/205
omissions	15/17/20	24/31/34	13/17/19
commissions	32/34/35	51/56/60	9/13/15
precision	0.859/0.847/0.841	0.829/0.810/0.797	0.959/0.941/0.932
recall	0.927/0.917/0.902	0.911/0.885/0.874	0.942/0.924/0.915
F_1-score	0.890/0.881/0.871	0.868/0.846/0.834	0.950/0.932/0.923

Table 4
Differences of contingencies and metrics in the accuracy of visible boundary determination and DSM segmentation for the test plots in images with $\approx 0°/45°/52°$ zenit angle.

	Deciduous	Mixed	Coniferous
image	N0125/R0110/R0105	N0124/R0108/R0104	N0122/R0107/R0102
total segments	0/0/−1	0/−2/−2	0/0/0
detections	−3/−5/−8	−1/−9/−12	0/−4/−6
omissions	+3/+5/+8	+1/+8/+11	0/+4/+6
commissions	+3/+5/+6	+1/+6/+10	0/+4/+6
precision	−0.010/−0.022/−0.028	−0.003/−0.022/−0.035	0.0/−0.018/−0.027
recall	−0.014/−0.024/−0.039	−0.004/−0.030/−0.041	0.0/−0.018/−0.027
F_1-score	−0.014/−0.023/−0.033	−0.004/−0.026/−0.038	0.0/−0.018/−0.027

was out of image bounds. The decreasing numbers of segments in the oblique cases of the deciduous and mixed area are indeed caused by occlusions, where at most two points were visible. The associated groundtruth points were not removed. Concerning factors one and two, it can be observed that the result in the nadir view cases is similar to that of the DSM segmentation. Additional omissions and commissions that occured in the oblique view cases are prevailingly caused by segments that miss the groundtruth points very closely.

In general, due to occlusion, the result of visible boundary determination can be fundamentally different from the DSM segmentation, not only in terms of altered segment boundaries, but also in terms of existence of a segment. So in order to apply the same strategy for the accuracy detection as for the DSM, it is necessary to determine the subset of visible CCs. But this is still not sufficient, since a groundtruth tree top
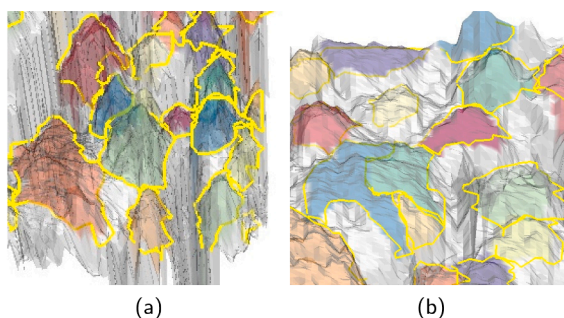
Fig. 12. Close up of the reprojected visible boundaries.

required. The question is whether the accuracy should be evaluated with respect to the image (which includes uncertainties from DSM generation) or with respect to the DSM (which enables a sensitivity analysis for α). In the former case, one groundtruth image is required for each image, which seems impractical for exhaustive evaluation and what is more, is that the delineation of coniferous trees in oblique views is precarious for human operators. In the latter case, no deterministic, qualitative statement about the alignment with the image objects can be made. Actually both cases need investigation to systematically account for errors. A practical solution will be outlined in Section 4.

As a consequence, pragmatically, reprojection of the boundary onto the mesh is performed for visual assessment. An example is given in Fig. 12 for a close-up part of the coniferous and deciduous region. Rendering of the reprojection is done without displaying the mesh edges, since backfaceculling needs to be turned off in the VTK graphics pipeline (cf. (Schroeder et al., 2006)) for the case when the reprojected boundary is decided to be below the surface.
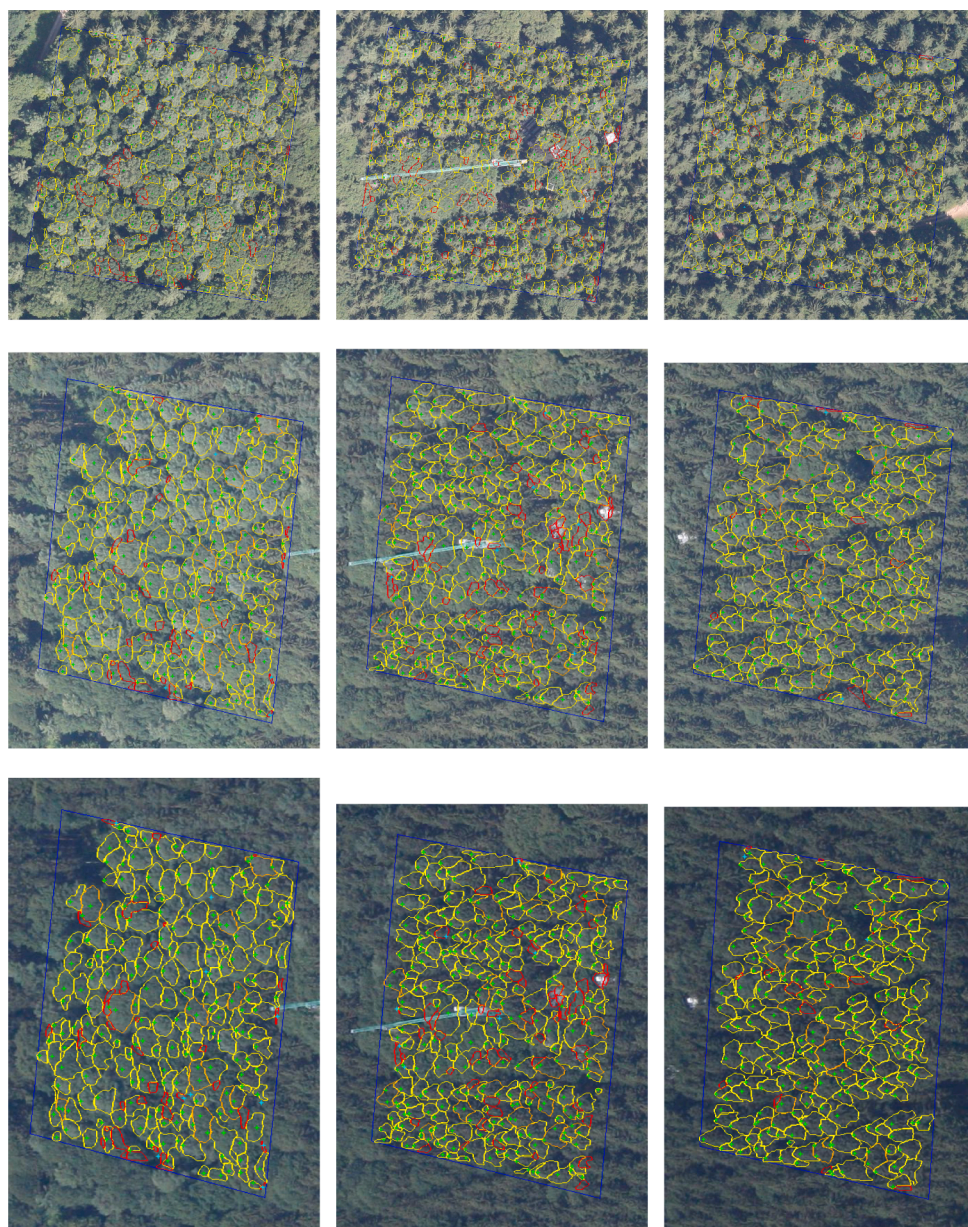
can be hidden, while parts of the tree are still visible. Apart from the implications from occlusion, the chosen α-value can substantially affect the result, as also demonstrated in Fig. 10.

It is concluded, that a region based groundtruth evaluation is

Fig. 13. Visible boundaries (displayed in yellow/orange/red for detection/omission/commission) and groundtruth tree tops (displayed as green/cyan markers for detection/omission) are overlaid on images of each plot (from left to right) in forward, nadir and backward view (from top to bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

The present approach avoids difficulties in image based correlation matching, introduced through analytic templates, by segmenting tree crowns in the DSM and transforming visible boundaries into the image. The labeling of a DSM segment is carried over to the image, also when it displays as disconnected parts in oblique views.

4.1. DSM segmentation

Referring to (Larsen et al., 2011), the proposed method is intended for dense, unmanaged forest types. Admittedly, the author believes that more recent extensions of ACWE allow to formulate individual tree crown segmentation solely as a global energy minimization, that does not require a detection or postprocessing step. The former seems prohibited, as deciduous trees do neither always inhibit a unique salient maxima, nor exhibit a typical boundary shape (especially when they are part of a tree cluster). This was the actual motivation to use the Chan-Vese segmentation. Critically, the piecewise constant Mumford-Shah image model is not appropriate for forests. The problem is threefold: (i) Trees have approximately the surface shape of an upper ellipsoid, such that the foreground objects are not constant. (ii) The height of different trees typically varies. (iii) The background can be either the ground between two trees or points above the ground, where canopies touch each other. Observation (i) implies that a piecewise smooth model is necessary, which can be solved by e.g. (Nielsen, 1997; Pock et al., 2009). (ii) and (iii) suggests that the segmentation has to consider the local context. However, the DSM is a topographic surface and so the topology of its decomposition into levelsets is known, which can constrain the segmentation model.

4.2. Visible boundary determination

Worst case and average case runtimes for building spatial data structures and performing intersection have been studied extensively in literature. They can differ significantly, depending on the point distribution, or in other words, on the topography and topology of the scene. As surfaces of densely forested areas are similar, it is worthwhile to evaluate the performance from an empirical and theoretical point of view in order to assure best performance, especially considering large datasets. In this respect, future investigations are planned on the use of triangle octrees (Bern and Plassmann, 2000) instead of OBB-trees. Moreover, the creation of the data structures as well as the intersection queries can be parallelized to accelerate runtime. Another notable alternative is the direct method of (Katz et al., 2007), where it is shown that the visible points of a point cloud are those, residing on the convex hull of its spherical flipping transform and the viewpoint.

The accuracy of the presented visibility detection is limited by the density of points: The vertices of a triangle can be visible while parts of the triangle are hidden. The maximum possible error of the resulting boundary effectively depends on the area of the triangle. For accurate boundaries, area intersection between the triangle and occluding projected triangles is required. A triangle A is occluding another triangle B if the line of sight through vertices of A intersects B. Thus, compared to the current implementation for visible points, also intersections behind a point (relating to the front to back order in the camera view) must be queried. The bespoke limitation can be alleviated through mesh refinement at the expense of runtime. Then, a triangle can be regarded as approximately visible, if all its vertices are visible. An in-depth analysis of the complexity of each scenario is out of the scope of the present discussion.

Following up, it should be mentioned that the boundary of visible points could be deduced directly as the exterior boundary of the cascaded union of visible triangles. This can be achieved in the 2.5D triangle mesh or in the image space. The latter requires to maintain the edge relations throughout the projection. However, there are some pitfalls with this approach: Visibility determination may remove points in the interior of a segment, depending on the topography of the segment itself and other segments in the line of sight. Therefore, the visible submesh can contain holes or fragment into several parts. This requires a rather tedious postprocessing step, just as the adjustment of the level of detail, by using e.g. spline fitting. In comparison, the α-shape readily handles all these issues.

Coming back to accuracy assessment, a practical way to describe the quality of boundaries can be achieved as follows: First, a region based groundtruth is created for the DSM. This reference, processed with ray casting and calculation of the α-shape according to a given image gives a contour, whose points can be associated with their standard deviation at the DSM location. The evaluation of the detected visible boundaries with respect to the visible reference can be accomplished as in (Brandtberg et al., 2003), but additionally considering the standard deviation of boundary points.

5. Conclusion

This contribution presented a novel approach to the problem of individual tree crown segmentation in oblique view aerial images of forested areas. The method was tested on images with zenit angles of approximately 0, 45 and 52 degrees for a deciduous, mixed and coniferous plot. Good detection performance with F_1-scores between 0.834 (for the mixed plot) to 0.923 (for the coniferous plot) were attained for oblique views with $\approx 52°$ zenit angle. For decreasing zenit angles, performance improved up to 0.868 (for the mixed plot) and 0.950 (for the coniferous plot), which almost equals the result for the DSM segmentation. The measured performance decrease for oblique views is largely attributed to accuracy assessment using simply transformed DSM groundtruth.

Since the method relies on the digital surface model, the accuracy of the final image boundaries is affected by height errors that propagate via the segmentation of the digital surface model and visible boundary determination w.r.t. oriented images. Different scenarios are conceivable for boundary refinement in the image, which is envisaged in future work, e.g. by respecting the standard deviation of the digital surface model in the definition of energy functionals. The asymmetric difference of visible segments and refined segments can further serve to iteratively optimize DSM generation, precisely the merging procedure of semi-global matching derived point clouds.

The principle of the method can be transferred to other oblique view delineation problems by adapting the segmentation method of the digital surface model to the objects in question. Moreover, replacing the camera position with the sun position allows to compute shadow regions for the direct illumination case. Most important for forest ecology, the proposed oblique view individual tree crown delineation permits to construct bidirectional reflection distribution functions with high accuracy and hence paves the way for high precision analyses of reflection effects.

6. Funding

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Calculation of the oblique visible ROI

Let a $RECT = (x_0, x_1, y_0, y_1)$ be the lateral rectangular extent of the $DSM : \Omega \rightarrow \mathbb{R}$, such as the three plots that were selected for segmentation tests. Further, let $LAB : \Omega \rightarrow Labels \cup \{0\}$ be the labeled DSM and let IMG be an oriented image. Simple projection of corner points, constructed from $RECT$ by introducing actual height coordinates of the DSM or fixed height coordinates leads to a quadrilateral that does not fit the visible projected segments. Specifically, if the height of far (near) side corners underestimates (overestimates) the far (near) side segment heights, then projection will be outside of the quadrilateral. Analogously, if the height of far (near) side corners overestimates (underestimates) the far (near) side segment heights, then projection will have an offset inside of the quadrilateral.

In order to obtain a ROI, that is similar to the quadrilateral constructed from the projection of corner points with fixed height, denoted Q_f, and bounds the extent of the union of segment polygons exactly, the following steps are carried out:

1. Calculate $Q_f = [q_0, q_1, q_2, q_3]$ with points in counterclockwise order.
2. Project the foreground points of DSM^{UTM} to IMG:

$$S = \{P(X, Y, Z) : (X, Y, Z) \in DSM^{UTM}|_{LAB \neq 0}\}.$$

3. Compute the convex hull of $S : H = CH(S)$.
4. Calculate the centroid c of Q_f.
5. For each side $l_i = (q_i, q_{(i+1) \bmod 4}), i = 0, \ldots, 3$ of Q_f:
   (a) For points r of H, such that $\overrightarrow{crq_i}$ is a rightturn and $\overrightarrow{crq_{(i+1) \bmod 4}}$ is a leftturn, store the signed distance between r and its projection onto l, r', as $DIST(r)$. The sign is defined positive, if $|\overrightarrow{cr}| < |\overrightarrow{cr'}|$ and negative otherwise.
   (b) Determine $r^* = argmax(DIST)$.
   (c) Perform parallel translation of l_i into r^* to obtain l_i'.
6. Calculate the intersections of consecutive l_i', i.e. the corners of the adapted quadrilateral Q_v.

Fig. 14 demonstrates the construction of the oblique visible ROI for forward looking images of the three test plots that were selected for the oblique view delineation experiment.
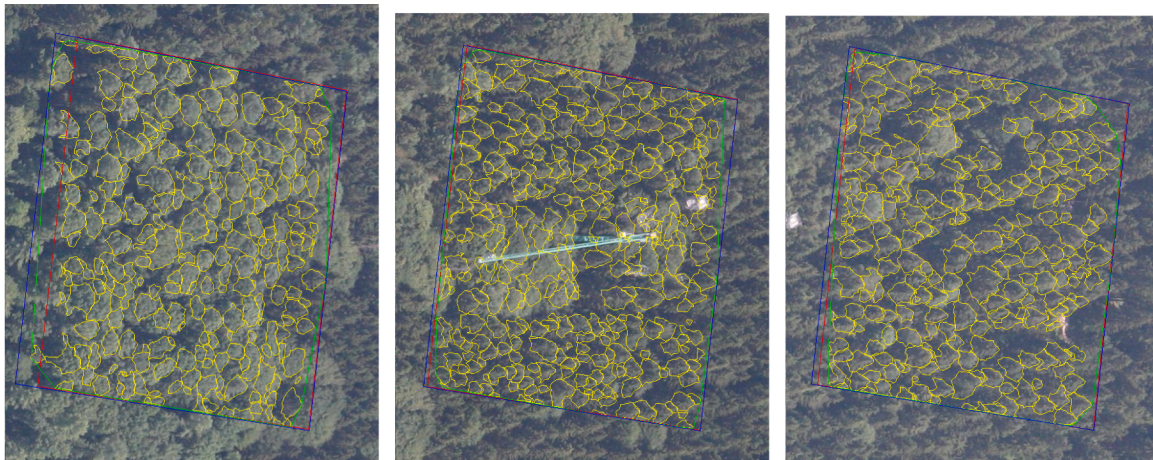

Fig. 14. The initial ROI, the convex hull of visible boundaries and the calculated visible ROI is shown in red, green and blue, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

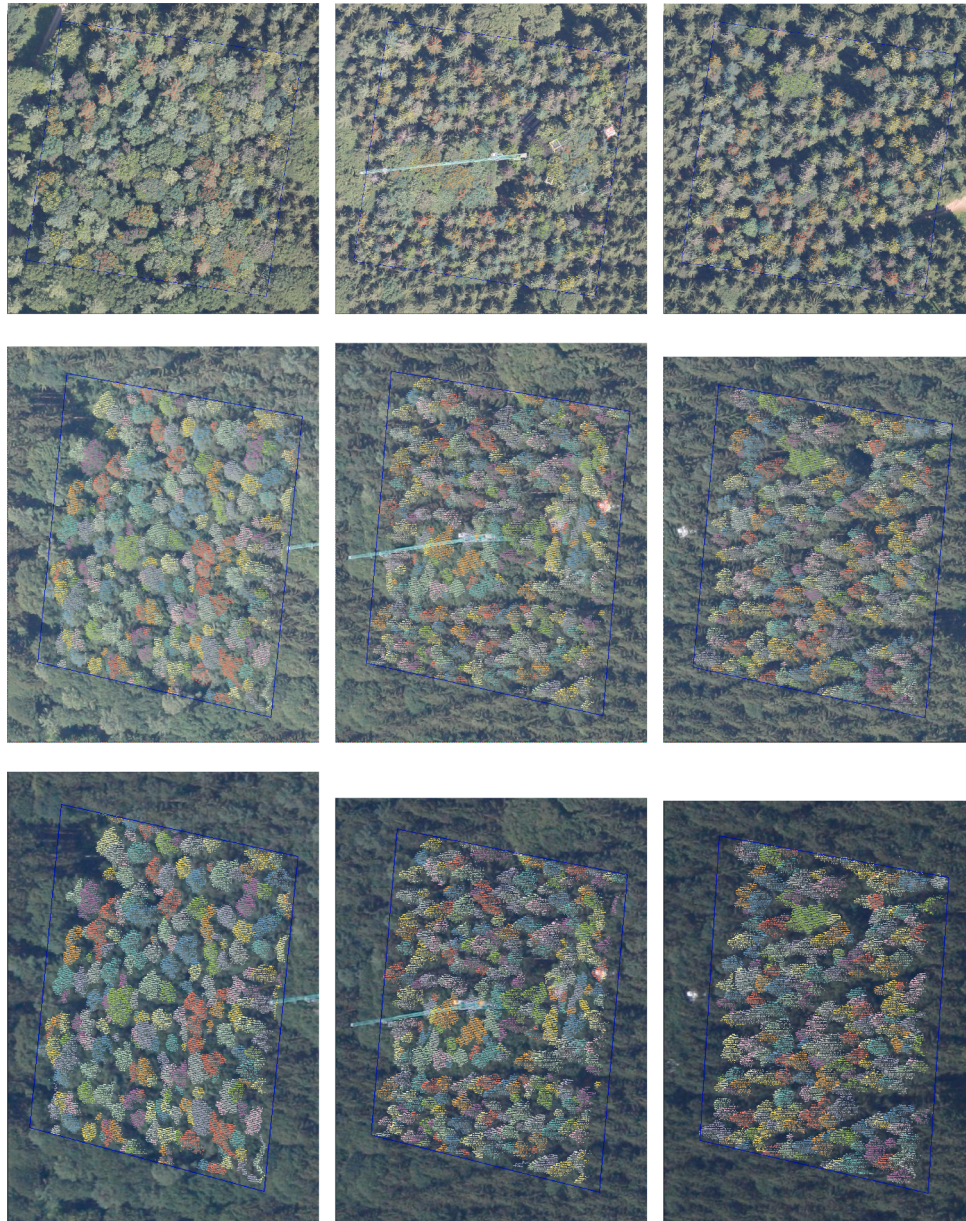## Appendix B. Results of visible point determination

See Fig. 15.

Fig. 15. Results of the visible point determination for each plot in forward, nadir and backward view.

References

Alsadik, B., Gerke, M., Vosselman, G., 2014. Visibility analysis of point cloud in close range photogrammetry. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial. Inf. Sci. II-5, 9–16. https://doi.org/10.5194/isprsannals-II-5-9-2014 https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-5/9/2014/.

Ardila, J.P., Bijker, W., Tolpekin, V.A., Stein, A., 2012. Multitemporal change detection of urban trees using localized region-based active contours in vhr images. Remote Sens. Environ. 124, 413–426. https://doi.org/10.1016/j.rse.2012.05.027 http://www.sciencedirect.com/science/article/pii/S0034425712002301.

Ardila, J.P., Tolpekin, V.A., Bijker, W., Stein, A., 2011. Markov-random-field-based super-resolution mapping for identification of urban trees in vhr images. ISPRS J. Photogram. Remote Sens. 66, 762–775.

Asaeedi, S., Didehvar, F., Mohades, A., 2013. Alpha convex hull, a generalization of convex hull. CoRR abs/1309.7829. http://arxiv.org/abs/1309.7829, arXiv: 1309.7829.

Asner, G.P., 2000. Contributions of multi-view angle remote sensing to land-surface and biogeochemical research. Remote Sens. Rev. 18, 137–162. https://doi.org/10.1080/02757250009532388.

Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7, 269–276. https://doi.org/10.1890/070152.

Baerentzen, J.A., Gravesen, J., Anton, F., Aans, H., 2012. Guide to Computational Geometry Processing: Foundations, Algorithms, and Methods. Springer Publishing Company Incorporated.

de Berg, M., Cheong, O., van Kreveld, M.J., Overmars, M.H., 2008. Computational geometry: algorithms and applications, third ed. Springer.

Bern, M., Plassmann, P., 2000. Chapter 6 - mesh generation. In: Sack, J.R., Urrutia, J. (Eds.), Handbook of Computational Geometry. North-Holland, Amsterdam, pp. 291–332. https://doi.org/10.1016/B978-044482537-7/50007-3 http://www.sciencedirect.com/science/article/pii/B9780444825377500073.

Beucher, S., Lantuéjoul, C., 1979. Use of watersheds in contour detection. In: International Workshop on Image Processing.

Brandtberg, T., Walter, F., 1998. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. Mach. Vision Appl.

Brandtberg, T., Warner, T.A., Landenberger, R.E., McGraw, J.B., 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in north america. Remote Sens. Environ. 85, 290–303.

Chan, T.F., Vese, L.A., 2001. Active contours without edges. Trans. Img. Proc. 10, 266–277. https://doi.org/10.1109/83.902291.

Chen, J., Liu, J., Leblanc, S., Lacaze, R., Roujean, J.L., 2003. Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. Remote Sens. Environ. 84, 516–525. https://doi.org/10.1016/S0034-4257(02)00150-5.

Chen, J.M., Leblanc, S.G., 1997. A four-scale bidirectional reflectance model based on canopy architecture. IEEE Trans. Geosci. Remote Sens. 35, 1316–1337.

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F., Lewis, S. L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in southeast asian forests from airborne laser scanning data. Remote Sens. Environ. 194, 77–88. https://doi.org/10.1016/j. rse.2017.03.017 http://www.sciencedirect.com/science/article/pii/ S0034425717301098.

Cremers, D., 2007. Nonlinear dynamical shape priors for level set segmentation, in. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

d'Angelo, P., 2016. Improving semi-global matching: Cost aggregation and confidence measure, in: ISPRS Congress 2016, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 299–304. http://elib.dlr.de/104197/.

Delaunay, B., 1934. Sur la sphere vide. Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk 7, 793–800.

Descombes, X., Pechersky, E., 2006. Tree Crown Extraction using a Three State Markov Random Field. Research Report RR-5982. INRIA. URL https://hal.inria.fr/inria-00097555.

Dubayah, R., Dozier, J., 1986. Orthographic terrain views using data derived from digital elevation models. Photogram. Eng. Remote Sens. 52, 509–518.

Edelsbrunner, H., 1992. Weighted Alpha Shapes. Technical Report. University of Illinois. Champaign, IL, USA.

Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 1983. On the shape of a set of points in the plane. IEEE Trans. Inf. Theor. 29, 551–559. https://doi.org/10.1109/ TIT.1983.1056714.

Edelsbrunner, H., Mücke, E.P., 1994. Three-dimensional alpha shapes. ACM Trans. Graph. 13, 43–72. https://doi.org/10.1145/174462.156635.

Elatawneh, A., Wallner, A., Straub, C., Schneider, T., Knoke, T., 2013. Tree Species Classification By Multiseasonal High Resolution Satellite Data, in: ESA Living Planet Symposium, p. 264.

Ericson, C., 2004. Real-Time Collision Detection. CRC Press Inc, Boca Raton, FL, USA.

Erikson, M., 2003. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. Can. J. For. Res. 33, 1557–1563. https://doi.org/10.1139/x03-062 arXiv:https://doi.org/10.1139/x03-062.

Escalante, B., 2012. Remote Sensing: Advanced Techniques and Platforms. IntechOpen. URL https://books.google.de/books?id=rbqZDwAAQBAJ.

Fairchild, M.D., Daoust, D.J.O., Peterson, J., Berns, R.S., 1990. Absolute reflectance factor calibration for goniospectrophotometry. Color Res. Appl. 15, 311–320. https://doi.org/10.1002/col.5080150605 https://www.onlinelibrary.wiley.com/ doi/abs/10.1002/col.5080150605.

Fawcett, T., 2006. An introduction to roc analysis. Pattern Recogn. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010 rOC Analysis in Pattern Recognition.

Ferreira, M., Féret, J.B., Grau, E., Gastellu-Etchegorry, J.P., Shimabukuro, Y., Souza Filho, C., 2018. Retrieving structural and chemical properties of individual tree crowns in a highly diverse tropical forest with 3d radiative transfer modeling and imaging spectroscopy. Remote Sens. Environ. 211 https://doi.org/10.1016/j. rse.2018.04.023.

Foley, F., Van, F., Van Dam, A., Feiner, S., Hughes, J., Angel, E., Hughes, J., 1996. Computer Graphics: Principles and Practice. Addison-Wesley systems programming series. Addison-Wesley.

Gastellu-Etchegorry, J.P., Lauret, N., Yin, T., Landier, L., Kallel, A., Malenovský, Z., Al Bitar, A., Aval, J., Benhmida, S., Qi, J., Medjdoub, G., Guilleux, J., Chavanon, E., Cook, B., Morton, D., Chrysoulakis, N., Mitraka, Z., 2017. Dart: Recent advances in remote sensing data modeling with atmosphere, polarization, and chlorophyll fluorescence. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing PP, 1–10. doi:10.1109/JSTARS.2017.2685528.

Gerke, M., Vosselman, G., 2016. Auswertung von Schrägluftaufnahmen. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 1–22. https://doi.org/10.1007/978-3-662-46900-2_57-1.

Getreuer, P., 2012. Chan-vese segmentation. Image Process. On Line 2, 214–224. https://doi.org/10.5201/ipol.2012.g-cv.

Gomes, M.F., Maillard, P., 2014. Integration of marked point processes and template matching for the identification of individual tree crowns in an urban and a wooded savanna environment in Brazil. In: Proc. SPIE 9245, Earth Resources and Environmental Remote Sensing/GIS Applications V, SPIE. doi:doi: 10.1117/12.2066848.

Gomes, M.F., Maillard, P., 2016. Detection of tree crowns in very high spatial resolution images. In: Marghany, M. (Ed.), Environmental Applications of Remote Sensing. InTech, Rijeka. chapter 02. doi:10.5772/62122, doi:10.5772/62122.

Gomes, M.F., Maillard, P., Deng, H., 2018. Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model. Remote Sens. Environ. 211, 184–195. https://doi.org/10.1016/j.rse.2018.04.002 http://www.sciencedirect.com/science/article/pii/S0034425718301470.

Gonzalez, R.C., Woods, R.E., 2001. Digital Image Processing, 2nd ed. Addison-Wesley Longman Publishing Co., Inc, Boston, MA, USA.

Gottschalk, S., Lin, M.C., Manocha, D., 1996. Obbtree: A hierarchical structure for rapid interference detection. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, NY, USA. pp. 171–180. URL http://doi.acm.org/10.1145/237170.237244, doi:10.1145/237170.237244.

Gottschalk, S.A., 2000. Collision Queries Using Oriented Bounding Boxes. Ph.D. thesis. The University of North Carolina. Chapel Hill. AAI9993311.

Guibas, L.J., Knuth, D.E., Sharir, M., 1992. Randomized incremental construction of delaunay and voronoi diagrams. Algorithmica 7, 381–413. https://doi.org/10.1007/ BF01758770.

Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, Second ed.,. Cambridge University Press.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341. https://doi.org/ 10.1109/TPAMI.2007.1166.

Hueni, A., Nieke, J., Schopfer, J., Kneubühler, M., Itten, K., 2009. The spectral database specchio for improved long-term usability and data sharing. Comput. Geosci. 35, 557–565. https://doi.org/10.1016/j.cageo.2008.03.015. URL http://www. sciencedirect.com/science/article/pii/S0098300408002720.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sens. Environ. 130, 245–253.

Ji, Z., Liao, Y., Zheng, L., Wu, L., Yu, M., Feng, Y., 2019. An assembled detector based on geometrical constraint for power component recognition. Sensors 19, 3517. https://doi.org/10.3390/s19163517.

Johnson, J.L., Padgett, M.L., 1999. Pcnn models and applications. IEEE Trans. Neural Networks 10, 480–498.

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., Dalponte, M., van Ewijk, K.Y., Forrester, D.I., Haeni, M., Higgins, S.I., Holdaway, R.J., Iida, Y., Lorimer, C., Marshall, P.L., Momo, S., Moncrieff, G.R., Ploton, P., Poorter, L., Rahman, K.A., Schlund, M., Sonké, B., Sterck, F.J., Trugman, A.T., Usoltsev, V.A., Vanderwel, M.C., Waldner, P., Wedeux, B.M.M., Wirth, C., Wöll, H., Woods, M., Xiang, W., Zimmermann, N.E., Coomes, D.A., 2017. Allometric equations for integrating remote sensing imagery into forest monitoring programmes. Glob. Change Biol. 23, 177–190. https://doi.org/10.1111/gcb.13388 http://arxiv.org/ abs/https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13388, arXiv:https:// onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.13388.

Katz, S., Tal, A., Basri, R., 2007. Direct visibility of point sets. ACM Trans. Graph. 26 https://doi.org/10.1145/1276377.1276407, 24–es.

Kempf, C., Tian, J., Kurz, F., d'Angelo, P., Reinartz, P., 2019. Local versus global variational approaches to enhance watershed transformation based individual tree crown segmentation of digital surface models from 3k optical imagery. ISPRS - Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. XLII-2/W13, 43–50. https://doi. org/10.5194/isprs-archives-XLII-2-W13-43-2019 https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W13/43/2019/.

Key, T., Warner, T.A., McGraw, J.B., Fajvan, M.A., 2001. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. Remote Sens. Environ. 75, 100–112. https://doi.org/10.1016/S0034-4257(00)00159-0 http://www. sciencedirect.com/science/article/pii/S0034425700001590.

Koch, B., 2010. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS J. Photogram. Remote Sens. 65, 581–590. https://doi.org/10.1016/j. isprsjprs.2010.09.001 http://www.sciencedirect.com/science/article/pii/ S0924271610000845, iSPRS Centenary Celebration Issue.

Kurz, F., 2009. Accuracy assessment of the dlr 3k camera system. In: DGPF Tagungsband 18 2009.

Kurz, F., Müller, R., Stephani, M., Reinartz, P., Schroeder, M., 2007. Calibration of a wide-angle digital camera system for near real time scenarios. In: ISPRS Hannover Workshop 2007, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 1682–1777.

Lankton, S., Tannenbaum, A., 2008. Localizing region-based active contours. IEEE Trans. Image Process. 17, 2029–2039.

Larsen, M., 1999. Individual tree crown position estimation by template voting. In: Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing. ERIM International, Ottawa, Ontario, Canada.

Larsen, M., Eriksson, M., Descombes, X., Perrin, G., Brandtberg, T., Gougeon, F.A., 2011. Comparison of six individual tree crown detection algorithms evaluated under varying forest conditions. Int. J. Remote Sens. 32, 5827–5852. https://doi.org/ 10.1080/01431161.2010.507790 arXiv:https://doi.org/10.1080/ 01431161.2010.507790.

Larsen, M., Rudemo, M., 1998. Optimizing templates for finding trees in aerial photographs. Pattern Recogn. Lett. 19, 1153–1162.

Leitloff, I., Rosenbaum, D., Kurz, F., Meynberg, O., Reinartz, P., 2014. An operational system for estimating road traffic information from aerial images. Remote Sens. 6, 11315–11341. https://doi.org/10.3390/rs61111315 http://www.mdpi.com/2072-4292/6/11/11315.

Li, Z., Hayward, R., Zhang, J., Liu, Y., Walker, R., 2009. Towards automatic tree crown detection and delineation in spectral feature space using pcnn and morphological reconstruction. In: 16th IEEE International Conference on Image Processing (ICIP), pp. 1705–1708. https://doi.org/10.1109/ICIP.2009.5413642.

Lindeberg, T., 1994. Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Norwell, MA, USA.

Liu, J., Skidmore, A.K., Jones, S., Wang, T., Heurich, M., Zhu, X., Shi, Y., 2018. Large off-nadir scan angle of airborne lidar can severely affect the estimates of forest structure metrics. ISPRS J. Photogram. Remote Sens. 136, 13–25. https://doi.org/10.1016/j.isprsjprs.2017.12.004.

Martin, R.E., Asner, G.P., Bentley, L.P., Shenkin, A., Salinas, N., Huaypar, K.Q., Pillco, M. M., Ccori Álvarez, F.D., Enquist, B.J., Diaz, S., Malhi, Y., 2020. Covariance of sun and shade leaf traits along a tropical forest elevation gradient. Front. Plant Sci. 10, 1810. https://doi.org/10.3389/fpls.2019.01810.

McKenna, M., 1987. Worst-case optimal hidden-surface removal. ACM Trans. Graph. 6, 19–28. https://doi.org/10.1145/27625.27627.

Moreira, A.J.C., Santos, M.Y., 2007. Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. In: Braz, J., Vázquez, P., Pereira, J.M. (Eds.), GRAPP 2007, Proceedings of the Second International Conference on Computer Graphics Theory and Applications, Barcelona, Spain, March 8–11, 2007, Volume GM/R, INSTICC - Institute for Systems and Technologies of Information, Control and Communication. pp. 61–68.

Nicodemus, F., Richmond, J., Hsia, J., Ginsberg, I., Limperis, T., 1977. Geometrical considerations and nomenclature for reflectance. NBS Monograph 160.

Nielsen, M., 1997. Graduated nonconvexity by functional focusing. IEEE Trans. Pattern Anal. Mach. Intell. 19, 521–525.

Ning, X., Wang, Y., Meng, W., Zhang, X., 2016. Optimized shape semantic graph representation for object understanding and recognition in point clouds. Opt. Eng. 55, 1–14. https://doi.org/10.1117/1.OE.55.10.103111.

Osher, S., Sethian, J.A., 1988. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. 79, 12–49.

Perrin, G., Descombes, X., Zerubia, J., 2005. Adaptive simulated annealing for energy minimization problem in a marked point process application. In: Rangarajan, A., Vemuri, B., Yuille, A.L. (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition: 5th International Workshop, EMMCVPR 2005, St. Augustine, FL, USA, November 9–11, 2005. Proceedings, Springer, Berlin Heidelberg, Berlin, Heidelberg. pp. 3–17. doi: 10.1007/11585978_1, doi:10.1007/11585978_1.

Perrin, G., Descombes, X., Zerubia, J., 2006. 2d and 3d vegetation resource parameters assessment using marked point processes. In: 18th International Conference on Pattern Recognition (ICPR'06), pp. 1–4. doi:10.1109/ICPR.2006.20.

Pieczynski, W., Tebbache, A.N., 2000. Pairwise markov random fields and segmentation of textured images. Mach. Graph. Vision 9 (3), 705–718.

Pock, T., Cremers, D., Bischof, H., Chambolle, A., 2009. An algorithm for minimizing the mumford-shah functional, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 1133–1140. https://doi.org/10.1109/ICCV.2009.5459348.

Polewski, P., Yao, W., Heurich, M., Krzystek, P., Stilla, U., 2015. Detection of single standing dead trees from aerial color infrared imagery by segmentation with shape and intensity priors. Pia15+Hrigi15 - Joint Isprs Conference, Vol. Ii, Commission W4, 181–188. URL http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-34586-9.

Pollock, R.J., 1994. Model-based approach to automatically locating tree crowns in high spatial resolution images, in: Proc.SPIE, p. 12. doi: 10.1117/12.196753, doi:10.1117/12.196753.

Pollock, R.J., 1996. The Automatic Recognition of Individual Trees in Aerial Images of Forests Based on a Synthetic Tree Crown Image Model. Ph.D. thesis. The University of British Columbia (Canada). AAINN14815.

Popescu, S., Wynne, R., Nelson, R., 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Canad. J. Remote Sens. 29, 564–577.

Pretzsch, H., Grams, T., Häberle, K., Pritsch, K., Bauerle, T., Rötzer, T., 2020. Growth and mortality of norway spruce and european beech in monospecific and mixed-species stands under natural episodic and experimentally extended drought. results of the kroof throughfall exclusion experiment. Trees, 1–14.

Remondino, F., Gerke, M., 2015. Oblique aerial imagery: a review. In: Frietsch, D. (Ed.), Proceedings of Photogrammetric Week '15, 7–11 September, Stuttgart, Germany, Wichmann. pp. 75–83.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.

Roosjen, P., Clevers, J., Bartholomeus, H., Schaepman, M., Schaepman-Strub, G., Jalink, H., van der Schoor, R., de Jong, A., 2012. A laboratory goniometer system for measuring reflectance and emittance anisotropy. Sensors 12, 17358–17371. https://doi.org/10.3390/s121217358.

Schaepman, M.E., 2007. Spectrodirectional remote sensing: From pixels to processes. Int. J. Appl. Earth Observ. Geoinform., 9, 204–223. http://www.sciencedirect.com/science/article/pii/S0303243406000444, doi: 10.1016/j.jag.2006.09.003. advances in airborne electromagnetics and remote sensing of agro-ecosystems.

Schroeder, W., Martin, K., Lorensen, B., Sobierajski, L., Avila, R., Law, C., 2006. Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, 4th. ed.,. Kitware.

Straub, B.M., 2003. Automatic extraction of trees from aerial images and surface models. ISPRS Archives XXXIV, Part 3/W8, Munich, 17.-19.9.2003.

Tharwat, A., 2018. Classification assessment methods. Appl. Comput. Inform. https://doi.org/10.1016/j.aci.2018.08.003.

Tolpekin, V., Ardila, J., Bijker, W., 2010. Super-resolution mapping for extraction of urban tree crown objects from vhr satellite images. In: Addink, E., van Boillie, F. (Eds.), GEOBIA 2010: geographic object - based image analysis, 29 June-2 July 2010, Ghent, Belgium: proceedings, International Society for Photogrammetry and Remote Sensing (ISPRS). pp. –.

Trochta, J., Krucek, M., Vrška, T., Král, K., 2017. 3d forest: An application for descriptions of three-dimensional forest structures using terrestrial lidar. PLOS ONE 12, 1–17. https://doi.org/10.1371/journal.pone.0176871.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imag. 22, 137–154.

Voronoi, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs. Journal für die reine und angewandte Mathematik 134, 198–287 http://eudml.org/doc/149291.

Wang, L., Gong, P., Biging, G.S., 2004. Individual tree crown delineation and tree-top detection in high spatial resolution aerial imagery. Photogram. Eng. Remote Sens. 70, Nr.3.

Ward, G.J., 1992. Measuring and modeling anisotropic reflection. In: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques. Association for Computing Machinery, New York, NY, USA, pp. 265–272. https://doi.org/10.1145/133994.134078.

White, J.C., Coops, N.C., Wulder, M.A., Vastaranta, M., Hilker, T., Tompalski, P., 2016. Remote sensing technologies for enhancing forest inventories: A review. Canad. J. Remote Sens. 42, 619–641. https://doi.org/10.1080/07038992.2016.1207484.

Whitted, T., 1980. An improved illumination model for shaded display. Commun. ACM 23, 343–349. https://doi.org/10.1145/358876.358882 http://doi.acm.org/10.1145/358876.358882.

Yan, Z., Liu, R., Cheng, L., Zhou, X., Ruan, X., Xiao, Y., 2019. A concave hull methodology for calculating the crown volume of individual trees based on vehicle-borne lidar data. Remote Sens. 11, 623. https://doi.org/10.3390/rs11060623.

Zhang, J., Sohn, G., 2010. A markov random field model for individual tree detection from airborne laser scanning data. IAPRS, Vol. XXXVIII, Part 3A.

Zhao, T., Yang, Y., Niu, H., Wang, D., Chen, Y., 2018. Comparing U-Net convolutional network with mask R-CNN in the performances of pomegranate tree canopy segmentation. In: Larar, A.M., Suzuki, M., Wang, J. (Eds.), Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques and Applications VII, International Society for Optics and Photonics. SPIE. pp. 210–218. doi: 10.1117/12.2325570, doi:10.1117/12.2325570.

Zhuo, X., Mönks, M., Esch, T., Reinartz, P., 2019. Facade segmentation from oblique uav imagery. In: 2019 Joint Urban Remote Sensing Event (JURSE). IEEE, pp. 1–4.

Christian Kempf received his Diploma (Dipl.Inf.) in computer science from the University of Jena in 2009, completing his thesis at IBM R&D, Böblingen. He worked as a freelance programmer for the Max Planck Insitute for Bio-Geo-Chemistry in Jena, as a research associate at the Department of 3D Data Processing, Society for the Promotion of applied Computer Science (GFaI e.V.) in Berlin and as software developer at the Faculty of Geoscience and Earth Observation, University of Twente in Enschede. Since 2017 he is pursuing his Ph.D., jointly at the Institute of Forest Management, Technical University of Munich in Weihenstephan and the Department of Photogrammetry and Image Analysis at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF) in Oberpfaffenhofen. His current research focus is on the application of photogrammetric, image processing and computer graphics methods for early detection of drought stress in aerial imagery.

Jiaojiao Tian received the B.S. degree in geoinformation systems from the China University of Geosciences, Beijing, China, in 2006, the M.Eng. degree in cartography and geoinformation from the Chinese Academy of Surveying and Mapping, Beijing, in 2009, and the Ph.D. degree in mathematics and computer science from Osnabrück University, Osnabrück, Germany, in 2013. Since 2009, she has been with the Photogrammetry and Image Analysis Department, Remote Sensing Technology Institute, German Aerospace Center, Weßling, Germany, where she is currently the Head of the 3D Modeling Team. In 2011, she was a Guest Scientist with the Institute of Photogrammetry and Remote Sensing, ETH Zürich, Zurich, Switzerland. Her research interests include 3-D change detection, digital surface model (DSM) generation and quality assessment, object extraction, and DSM-assisted building reconstruction and classification.

Franz Kurz received the Ph.D. degree in engineering science from the Technical University Munich in 2003. He held post-doctoral positions at the Chair for Photogrammetry and Remote Sensing at Technical University Munich and later at the Cartographic Institute of Catalonia in Barcelona. Since 2005, he has been with German Aerospace Center (DLR) at the Remote Sensing Technology Institute in Oberpfaffenhofen. As a project manager and researcher, he has been involved in different national and European projects with focus on remote sensing applications. His main research interests are methods for real-time photogrammetry and traffic related applications in the field of remote sensing. Since 2018, he is co-team leader of the Hyperspectral Remote Sensing and Traffic Monitoring group at the Remote Sensing Technology Institute.

Pablo d'Angelo received his diploma degree in computer engineering (Dipl.-Ing FH) from the University of Applied Sciences Ulm in 2004. He has worked with Daimler AG from 2004 to 2007 on industrial computer vision and received his Ph.D. (Dr.-Ing.) in computer science from Bielefeld University in 2007 with a dissertation on joint use of geometric and photometricmethods for 3D reconstruction from optical images. In 2007, he joined the Photogrammetry and Image Analysis department of the Remote Sensing Technology Institute at the German Aerospace Center (DLR) in Oberpfaffenhofen. His main research topic are 3D reconstruction from remotely sensed stereo imagery, with a focus on operational systems for large scale image orientation and generation of digital elevation models.

Thomas Schneider received his Diploma in Geology and his PhD in Forest Sciences from LMU Munich in 1985 and 1994, respectively. He is currently assistant professor at the Institute of Forest Management, heading the Remote Sensing group, and at Limnological Station of the Technical University of Munich. His research interest is focused on field and laboratory spectroscopy, goniometry and BRDF approximation in forestry, agriculture and limnology. Beyond that, he worked in remote sensing and GIS based expert system developmentfor these application domains. He has been project scientist and project leader in numerous national and international projects and reviewer for established journals.

Peter Reinartz received his Diploma (Dipl.-Phys.) in theoretical physics in 1983 from the University of Munich and his PhD (Dr.-Ing) in civil engineering from the University of Hannover, in 1989. His dissertation is on optimization of classification methods for multispectral image data. Currently he is department head of the department "Photogrammetry and Image Analysis" at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF) and holds a professorship for computer science at the University of Osnabrueck. He has more than 30 years of experience in image processing and remote sensing and over 400 publications in these fields. His main interests are in machine learning, stereo-photogrammetry and data fusion using space borne and airborne image data, generation of digital elevation models and interpretation of very high resolution data from sensors like WorldView, GeoEye, and Pleiades. He is also engaged in using remote sensing data for disaster management and using high frequency time series of airborne image data for real time image processing and for operational use in case of disasters as well as for traffic monitoring.