

DETECTION OF SINGLE STANDING DEAD TREES FROM AERIAL COLOR INFRARED IMAGERY BY SEGMENTATION WITH SHAPE AND INTENSITY PRIORS

P. Polewski^{a, b, *}, W. Yao^a, M. Heurich^c, P. Krzystek^a, U. Stilla^b

^a Dept. of Geoinformatics, Munich University of Applied Sciences, 80333 Munich, Germany -
(polewski, yao, krzystek)@hm.edu

^b Photogrammetry and Remote Sensing, Technische Universität München, 80333 Munich, Germany - stilla@tum.de

^c Dept. for Research and Documentation, Bavarian Forest National Park, 94481 Grafenau, Germany -
marco.heurich@npv-bw.bayern.de

KEY WORDS: Segmentation, level sets, active contours, dead tree detection, aerial CIR imagery

ABSTRACT:

Standing dead trees, known as snags, are an essential factor in maintaining biodiversity in forest ecosystems. Combined with their role as carbon sinks, this makes for a compelling reason to study their spatial distribution. This paper presents an integrated method to detect and delineate individual dead tree crowns from color infrared aerial imagery. Our approach consists of two steps which incorporate statistical information about prior distributions of both the image intensities and the shapes of the target objects. In the first step, we perform a Gaussian Mixture Model clustering in the pixel color space with priors on the cluster means, obtaining up to 3 components corresponding to dead trees, living trees, and shadows. We then refine the dead tree regions using a level set segmentation method enriched with a generative model of the dead trees' shape distribution as well as a discriminative model of their pixel intensity distribution. The iterative application of the statistical shape template yields the set of delineated dead crowns. The prior information enforces the consistency of the template's shape variation with the shape manifold defined by manually labeled training examples, which makes it possible to separate crowns located in close proximity and prevents the formation of large crown clusters. Also, the statistical information built into the segmentation gives rise to an implicit detection scheme, because the shape template evolves towards an empty contour if not enough evidence for the object is present in the image. We test our method on 3 sample plots from the Bavarian Forest National Park with reference data obtained by manually marking individual dead tree polygons in the images. Our results are scenario-dependent and range from a correctness/completeness of 0.71/0.81 up to 0.77/1, with an average center-of-gravity displacement of 3-5 pixels between the detected and reference polygons.

1. INTRODUCTION

Dead wood, both downed and standing, plays an important part in forest biodiversity by providing habitat for plants and animals (Fridman and Walheim, 2000). Also, it forms a significant component of forest carbon stocks (Woodall et al., 2008). Furthermore, snags are a source of coarse woody debris which is important in stand succession (Harmon et al., 1986). Therefore, obtaining information about the distribution of snags is necessary in biodiversity and carbon dynamics studies as well as in forest management tasks. Another important reason which drives the need for snag detection techniques is the investigation of disease and insect infestation patterns in forests (Lausch et al., 2013).

The use of infrared bands for assessing vegetation health based on aerial and satellite imagery is quite widespread (e.g. Wang et al. (2007), Vogelmann (1990)). Both the intensity of the infrared channels and various indices derived from it, e.g. NDVI (Tucker, 1979), are useful for distinguishing between dead wood and living vegetation due to the reflectance differences in these spectral bands (Jensen, 2006). Several authors have developed manual or semi-automatic individual snag detection methods which make use of this fact (Bütler and Schlaepfer (2004), Heurich et al. (2010), Pasher and King (2009)). The disadvantage of these approaches is the necessity to involve a human expert in the processing, which may limit their applicability to large areas. Another group of methods aims at estimating the total biomass of snags at the plot or raster level. This can be done both based on ALS point clouds (Kim et al., 2009) and satellite/aerial imagery (Eskelson et al., 2012). Finally, there seem to be relatively few

contributions providing a fully automatic methodology for detecting single dead trees at the object level. Yao et al. (2012) first perform single tree segmentation in ALS point clouds and subsequently classify each tree segment as either dead or living based on geometric properties of the segment as well as radiometric features of the laser points. They report an overall classification accuracy of 71-73% in the leaf-off and leaf-on state, respectively. Lee and Cho (2006) apply a local pixel intensity maxima filtering algorithm to isolate individual tree crowns and investigate the use of 4 spectral (RGB + near infrared) bands for discriminating between infested and healthy trees. They note that the quality of the crown detection is sensitive to the correct choice of the window size parameter for filtering. More recently, Bhattarai et al. (2012) adopt a similar strategy for classifying *Sirex*-infested pines from 8-band WorldView-2 imagery. They first perform delineation of individual tree crowns using the algorithm by Ke et al. (2010), which involves active contour segmentation and subsequent refinement using template matching and heuristic rules. Each tree crown is then classified as either infested or healthy based on features derived from the 8 spectral bands. The overall classification accuracy at the object level attains 0.81, however the authors observe that the delineation algorithm failed to find a significant portion of the smaller crowns and point out that this part of their method has improvement potential.

The three aforementioned snag detection methods all share the characteristic of first attempting to segment the input data into individual trees, without regard for whether they are dead or living, and only perform the classification in the second step. In this paper, we propose an alternate strategy where the segmentation process itself is geared towards the traits of dead trees. We em-

*Corresponding author.

ploy a level set segmentation scheme which evolves the image region associated with a candidate dead tree in a way consistent, in terms of both shape and pixel intensities, with prior information derived from external training data. Based on the results of numerous studies in other applications (e.g. Ye et al. (2010), Tsai et al. (2003), Cremers et al. (2006)), we argue that the introduction of object-specific priors can benefit the segmentation process also for the task of snag detection. Moreover, the prior information alters the energy functional of the segmentation in such a manner that it will evolve towards an empty contour in the absence of the actual target object in the relevant image region. This leads to an implicit detection scheme and eliminates the need for a separate classification step.

Over the last two decades, level set methods have emerged as one of the most prominent frameworks for image segmentation. The key idea is to represent the evolving object contour curve $C \in \Omega$, where $\Omega \subset R^2$ is the image domain, as the zero level set of an embedding function $\phi : \Omega \rightarrow R$. In such a setting, the evolution of C is done implicitly through the direct evolution of ϕ . Among the most important advantages of this modification, comparing to a representation of C via a set of control points, is the intrinsic handling of object topology changes (i.e. splitting into several parts) and avoiding the necessity to check for self-intersection of the contour and to perform elaborate re-gridding (Cremers et al., 2007). Level set methods have been successfully applied to region-based segmentation schemes, wherein the optimized cost functional depends on the properties of entire regions into which the image is partitioned by the contour C . Region-based approaches constituted a major improvement over purely edge-based methods in that they are less sensitive to noise and possess less local minima in their associated energy functionals. During the last 10 years, considerable research effort has been directed towards incorporating shape and intensity prior information into the segmentation framework. For this study, we chose to adopt the method proposed by Cremers and Rousson (2007), which is data-driven in the sense that it uses non-parametric statistical models for both shape and intensity priors. Since this method was originally introduced in the context of segmenting medical images with a single intensity channel, we extended the intensity model to account for the 3 color channels present in our scenario.

The rest of this paper is structured as follows: in Section 2. we explain the details of our approach including the mathematical background. Section 3. describes the study area, experiments and evaluation strategy. The results are presented and discussed in Section 4. Finally, the conclusions are stated in Section 5.

2. METHODOLOGY

2.1 Overview of strategy

Our method for detecting individual snags accepts single color infrared images as input. The number of spectral bands is not fixed, however the assumption is that the image contains bands such that the joint distributions (over all bands) of values of pixels belonging to dead trees versus pixels belonging to other objects are significantly different. No image correspondence or 3D elevation information is necessary, but the image scale (pixel resolution at ground) is assumed to be known. Furthermore, we require that a set of training samples be available in the form of images with polygons delineating individual snags. The training data is the basis for deriving shape and intensity prior information. The entire algorithm can be conceptually divided into two major steps. The first step uses the intensity prior information to find regions

in the image which are likely to contain snags. In the second step, a fine segmentation is iteratively applied locally at different points of the image which cover the high-likelihood snag regions. This fine segmentation uses both intensity and shape priors. The output of our algorithm is a set of Boolean pixel masks (or equivalently, contours) which specify the sets of pixels of the original image that define each individual detected snag. An overview of the detection pipeline is shown in Fig. 1.

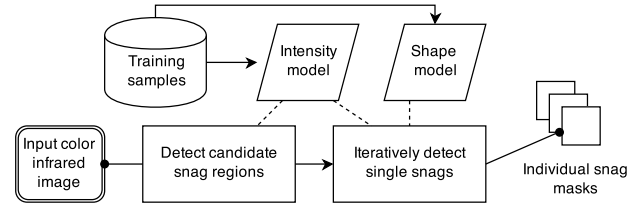


Figure 1. Overview of snag detection strategy.

2.2 Candidate region detection

This stage of the algorithm aims at finding all regions in the input image having a pixel intensity profile similar to that of dead trees. For this purpose, we perform a clustering of the input image using a Gaussian mixture model (GMM) with exactly 3 components, corresponding to the 3 types of areas commonly found in our target scenarios: (i) living vegetation, (ii) dead trees, and (iii) shadows. We chose the GMM due to its ability to model interactions between the pixel features via the covariance matrix, and also because of a statistically sound way to introduce prior information. The GMM is defined in a Bayesian setting, with priors on the component means and variances. Let $X = (x_1, x_2, \dots, x_N)$ denote the d -dimensional feature vectors of the N image pixels, and β_i, μ_i, Σ_i , respectively, the mixture coefficient, mean and covariance matrix of component $i, i = 1..3$. The posterior probability of the mixture parameters conditioned on the data is proportional to the product of the data likelihood \mathcal{L} and the prior parameter probability \mathcal{P}_p (Eq.1):

$$\mathcal{P}(\mu, \Sigma, \beta | X, \Theta) \propto \mathcal{L}(X | \mu, \Sigma, \beta) \mathcal{P}_p(\mu, \Sigma | \Theta) \quad (1)$$

In the above, we assume a flat prior on the mixture coefficients. The vector $\Theta = (\theta_1, \theta_2, \theta_3)$ represents hyper-parameters of the per-component prior distributions. The data likelihood is the standard probability of a multivariate GMM (Eq.2):

$$\mathcal{L}(X | \mu, \Sigma, \beta) = \prod_{j=1}^N \sum_{k=1}^3 \beta_k \mathcal{P}_k(x_j | \mu_k, \Sigma_k) \quad (2)$$

$$\mathcal{P}_k(x | \mu, \Sigma) = |2\pi\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\}$$

Following Fraley and Raftery (2007), we apply an inverse Wishart prior on the covariance matrices and a multivariate normal prior for the component means (Eq.3).

$$\mathcal{P}(\Sigma | \theta) \propto |\Sigma|^{-\left(\frac{\nu_p}{2} + 2\right)} \exp\left\{-\frac{1}{2} \text{trace}(\Sigma^{-1} \Lambda_p^{-1})\right\} \quad (3)$$

$$\mu_k | \Sigma_k, \theta \sim \mathcal{N}(\mu_{p,k}, \Sigma_k / \kappa_p)$$

In the above, the prior probability model for the covariance matrix is common, whereas the mean priors are component-specific. The hyper-parameter vector θ_k consists of four entries: $\theta_k = (\mu_{p,k}, \Lambda_p, \nu_p, \kappa_p)$. The prior means $\mu_{p,k}$ are calculated based on samples from appropriate patches in the training images. The rest of the hyper-parameters receive the same values for all components. The shrinkage κ_p is set to the product of N and a small positive constant, while the degrees of freedom parameter ν_p as-

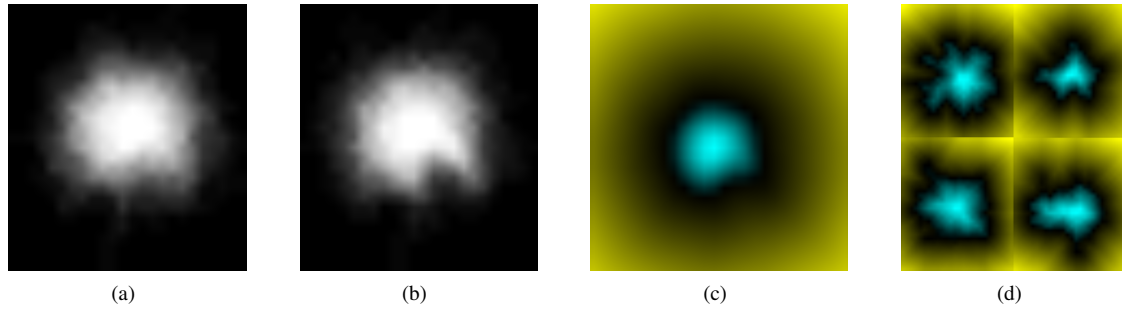


Figure 2. Deriving the eigenshape representation. (a), (b): mean polygon masks formed by summing all pixel values at corresponding locations in training polygons before and after alignment, respectively. (c): the mean signed distance function (SDF) of the training set. (d) SDFs of sample training shapes represented using 20 first eigenmodes of shape variation. In (c)-(d), the cyan and yellow colors represent the interior and exterior of the shape, respectively, with intensity proportional to the distance from the shape contour.

sumes a value of $d + 2$. The scale Λ_p is taken to be the covariance matrix of all data points X . Note that the prior on the covariance matrices acts only as a Bayesian regularizer of the parameter estimation problem and does not actually introduce prior information depending on the training examples. The prior distribution imposed on the means is of greater importance, as it attracts the cluster centers towards partitions with regions that roughly correspond to our three semantically different area types. The optimal parameters μ, Σ, β maximizing the posterior probability (Eq. 1) are found via the Expectation-Maximization (EM) algorithm, using the prior means and aggregate covariance matrix as the starting point (Fraleigh and Raftery, 2007). Despite the regularization, we found that the covariances can degenerate into singular matrices under certain conditions. Specifically, if the image contains no pixels whose values lie close enough to the prior mean $\mu_{p,i}$, then component i degenerates to a singular covariance and zero mixing coefficient. However, this is not a problem, since this situation can be easily detected. If the degenerated component corresponds to the dead tree prior, we take this as evidence that the image does not contain snags and stop the processing pipeline. Otherwise, we just discard the invalid components and proceed normally. The parameters calculated using the EM algorithm are then used to assign each input image pixel to its maximum-probability component. As a result of this step, we obtain the set of locations of pixels from the input image which were assigned to the component corresponding to our snag prior intensity mean. For an example, see the partition of Fig. 3(b) into the red, black, and gray regions. Note that since the presence or absence of a tree at a location will ultimately be decided by the fine segmentation, false positive pixels in the region step are not a serious problem. However, false negative pixels may lead to an irrecoverable error of missing a dead tree.

2.3 Level set segmentation with priors

2.3.1 General formulation Let $\Omega \subset R^2$ be the image plane, $I : \Omega \rightarrow R^d$ a vector-valued image, and C an evolving contour in the image I . To find an optimal segmentation of I with respect to an energy functional $E(C)$, we evolve the contour C in the direction of the negative energy gradient (Cremers et al., 2007). In the level set method, C is represented implicitly as the zero level set of an embedding function $\phi : \Omega \rightarrow R$. We then evolve ϕ according to an appropriate partial differential equation dependent on our energy functional E (Eq.4):

$$\frac{\partial \phi}{\partial t} = - \frac{\partial E}{\partial \phi} \quad (4)$$

A common choice for ϕ is the signed distance function, whose value for a point $p \in \Omega$ is the negative distance from p to the contour C if p is inside the contour, the positive distance from p

to C if p is outside C , and 0 if p lies on the contour itself.

2.3.2 Relationship to region-based segmentation Consider the problem of binary segmentation, i.e. partitioning the image I into two disjoint (not necessarily connected) regions Ω_1 and Ω_2 . Using the Bayesian rule, we can express the conditional probability of the partition given the image as:

$$\mathcal{P}(\Omega_1, \Omega_2 | I) \propto \mathcal{P}(I | \Omega_1, \Omega_2) \mathcal{P}(\Omega_1, \Omega_2) \quad (5)$$

Moreover, since the embedding function ϕ uniquely defines the contour C and hence the partition Ω_1, Ω_2 , we have that:

$$\begin{aligned} E(\phi) &= -\log(\mathcal{P}(\phi | I)) = -\log(\mathcal{P}(I | \phi)) - \log(\mathcal{P}(\phi)) \\ &= E_{img} + E_{shp} \\ E_{img} &= -\log(\mathcal{P}(I | \phi)), E_{shp} = -\log(\mathcal{P}(\phi)) \end{aligned} \quad (6)$$

This forms a decomposition of the partition energy $E(\phi)$ into the sum of an image term E_{img} and a shape term E_{shp} . If we further assume that region labellings are uncorrelated, i.e. $\mathcal{P}(I | \Omega_1, \Omega_2) = \mathcal{P}(I | \Omega_1) \mathcal{P}(I | \Omega_2)$, and also that image values inside a region are realizations of independent and identically distributed random variables (Cremers et al., 2007), we can write the image term as:

$$\begin{aligned} E_{img}(\phi) &= - \int_{\Omega_1} \log f_1(I(x)) dx - \int_{\Omega_2} \log f_2(I(x)) dx \\ &= - \int_{\Omega} H[\phi(x)] \log \frac{f_1(I(x))}{f_2(I(x))} dx + Z \end{aligned} \quad (7)$$

In Eq.7, f_1 and f_2 represent, respectively, the probability density functions of the image values in region Ω_1 and Ω_2 , $H(z)$ indicates the Heaviside function, while Z is a constant independent of ϕ . The models f_1, f_2 can either be specified a priori, or are assumed to follow a distribution from a known family but with unknown parameters, which are functions of the partition and must be co-estimated in the optimization process. Optimizing the functional in Eq. 6 with respect to ϕ is an infinite-dimensional problem which can be solved using calculus of variations.

2.3.3 Eigenshape representation Instead of directly optimizing the energy $E(\phi)$ in a variational framework, we take the approach described in the pioneering works of Leventon et al. (2000) and Tsai et al. (2003), where the signed distance function ϕ is constrained to a parametric representation based on the eigenmodes of a set of training shapes. Let $\phi_1, \phi_2, \dots, \phi_M$ be M signed distance functions of training shapes sampled on a discrete grid. We start by aligning the training samples with respect to translation and rotation using the gradient descent method given in Tsai et al. (2003) (Figs. 2(a), 2(b)). The mean training shape is denoted by $\bar{\phi}$ (Fig.2(c)). The shape variability matrix S is defined

as:

$$S = (\text{vec}(\phi_1 - \bar{\phi}) \quad \text{vec}(\phi_2 - \bar{\phi}) \quad \cdots \quad \text{vec}(\phi_M - \bar{\phi})) \quad (8)$$

In Eq.8, $\text{vec}(M)$ indicates an operator that creates a column vector out of a matrix M by stacking its columns. We subsequently calculate the eigenvectors ψ_1, \dots, ψ_c of the matrix SS^T corresponding to its largest c eigenvalues. The parametric representation of the signed distance function can be then expressed as:

$$\phi_{\alpha, h, \theta}(x) = \bar{\phi}(R_\theta x + h) + \sum_{i=1}^c \alpha_i \psi_i(R_\theta x + h) \quad (9)$$

Fig. 2(d) depicts the eigenmode representation of example training shapes. The parameter vector α models shape deformations, while the rotation angle θ and translation offset h define the rigid transformation of the training shapes. Similarly to Cremers and Rousson (2007), we omit scaling from the shape representation, because the scale of our target objects is directly represented by the training samples. The main benefit of representing ϕ in terms of a set of parameters is that we can replace the infinite-dimensional variational problem of optimizing the energy functional (Eq. 6) with respect to ϕ by a finite-dimensional optimization problem with respect to the parameters, carried out in the constrained subspace spanned by the training examples:

$$E(\phi) = E(\alpha, h, \theta) = -\log(\mathcal{P}(I|\alpha, h, \theta)) - \log(\mathcal{P}(\alpha, h, \theta)) \quad (10)$$

2.3.4 Prior information We are interested in deriving a shape prior $\mathcal{P}(\phi) = \mathcal{P}(\alpha, h, \theta)$ (Eq. 6) which models the variability of the training polygons' shapes. The statistical nature of the model ensures that the evolved contours remain similar to, but not necessarily the same as, our training shapes. For this purpose, we adopt the nonparametric approach outlined in (Cremers and Rousson, 2007). Note that we model only the shape variability parameters α and set a flat prior for the rigid transformation parameters θ, h , because we consider all rotation angles and offsets as equally likely to appear in our scenario. In accordance with the original method, we first calculate the representations $\alpha_1, \dots, \alpha_M$ of the M original training shapes ϕ_1, \dots, ϕ_M in terms of the eigenmodes ψ_1, \dots, ψ_c . The probability of any shape parameter vector α can now be expressed using a kernel density estimate with respect to the training shapes:

$$\mathcal{P}(\alpha) = \frac{1}{M\sigma^c} \sum_{i=1}^M K\left(\frac{\alpha - \alpha_i}{\sigma}\right), K(u) = \frac{1}{2\pi^{c/2}} \exp\left(-\frac{u^2}{2}\right) \quad (11)$$

For simplicity, the kernel bandwidth σ is set to the average nearest neighbor distance within the training set. Other, theoretically well-founded methods of estimating an optimal σ value are also available. The nonparametric shape prior has the advantage of not making too restrictive assumptions about the distribution of the shape parameters. Also, in comparison to the Gaussian or uniform model, it is able to better account for the fact that the space of signed distance functions is not linear. This is because it prefers shape vectors lying in the neighborhood of the training examples, which are known to represent valid signed distance functions.

The intensity prior complements the shape model by encouraging the evolving shape to include patches of pixels with values frequently occurring in the training polygons, and simultaneously discouraging regions whose pixels values are uncharacteristic of the target class. In our setting, this amounts to fixing the intensity models f_1, f_2 (Eq. 7) for the foreground and background regions. In particular, Cremers and Rousson (2007) proposed to

construct two independent kernel density estimators for this task. However, since their method was designed to accept images with a single intensity channel, we choose to employ a different modeling technique which is able to capture the inter-dependencies between the various channels present in our target images. To achieve this, we first observe that the image term E_{img} of the energy functional (Eq.7) does not require the knowledge of the individual per-region probabilities $f_1(I(x)), f_2(I(x))$ for any point x , but only their ratio $r = f_1(I(x))/f_2(I(x))$. Note that $f_1(y)$ and $f_2(y)$ represent the region label-conditional probabilities of observing a feature value y . We can represent the ratio r using inverted models which quantify the probability of point x having a label $l \in 1, 2$ given its feature value $I(x)$:

$$r(x) = \frac{f_1(I(x))}{f_2(I(x))} = \frac{\mathcal{P}(I(x)|l=1)}{\mathcal{P}(I(x)|l=2)} = \frac{\mathcal{P}(l=1|I(x))\mathcal{P}(I(x))}{\mathcal{P}(l=1)} \times \frac{\mathcal{P}(l=2)}{\mathcal{P}(l=2|I(x))\mathcal{P}(I(x))} = \frac{\mathcal{P}(l=1|I(x))\mathcal{P}(l=2)}{\mathcal{P}(l=2|I(x))\mathcal{P}(l=1)} \quad (12)$$

Therefore, r is a quotient of the ratio of posterior region label probabilities and the ratio of the prior region label probabilities. Moreover, since we are only interested in the logarithm of $r(x)$, we see that $\log(r(x)) = \frac{\mathcal{P}(l=1|I(x))}{\mathcal{P}(l=2|I(x))} - D$, where D is a constant, because the prior probabilities do not depend on the point x . The re-expressing of r in terms of posterior label probabilities adds a D -multiple of ϕ 's interior area to the energy E_{img} . In particular, for our choice of equal prior label probabilities, D is zero and hence the energy is not affected. This means that we can incorporate any binary discriminative model of the boundary between the foreground and background region into E_{img} . In this setting, we have that $\mathcal{P}(l=2|I(x)) = 1 - \mathcal{P}(l=1|I(x))$, so E_{img} can be rewritten as:

$$E_{img}(\phi) = - \int_{\Omega} H[\phi(x)] \log \frac{\mathcal{P}(l=1|I(x))}{1 - \mathcal{P}(l=1|I(x))} dx \quad (13)$$

The quantity $\log\{\mathcal{P}(l=1|I(x))/(1 - \mathcal{P}(l=1|I(x)))\}$ is reminiscent of the logit function. Indeed, if logistic regression is chosen as the discriminative model, then E_{img} boils down to:

$$E_{img}(\phi) = - \int_{\Omega} H[\phi(x)] \omega I(x) dx \quad (14)$$

In the above, the vector ω represents the trained coefficients of the logistic regression model. In order to retain the power of nonparametric approaches, we kernelize the logistic regression model by transforming the image value vector $I(x)$ into a new feature vector $F(x) = [K(\frac{I(x)-I(y_1)}{\sigma_k}), \dots, K(\frac{I(x)-I(y_T)}{\sigma_k})]$. The entries of $F(x)$ quantify the distances of the point x to the T training points y_i in the original image feature space measured by the kernel K , where the bandwidth σ_k is chosen using n -fold cross validation. The final form of the logit is then $\omega F(x)$. The introduction of intensity priors causes the segmentation energy functional to depend on the pixel values $I(x)$ only through their foreground probabilities $f_1(I(x))$, which means that the algorithm is effectively working on a one-channel probability image (see Fig. 3(c)).

Final energy formulation Having defined both the shape and image terms in the energy functional (Eq. 6), we can now write down the final form of the energy used in our approach:

$$E(\alpha, h, \theta) = - \int_{\Omega} H[\phi_{\alpha, h, \theta}(x)] \omega F(x) dx - \log \sum_{i=1}^M K\left(\frac{\alpha - \alpha_i}{\sigma}\right) \quad (15)$$

We set up the partial differential equations governing the evolution of the function ϕ in the direction of the negative energy

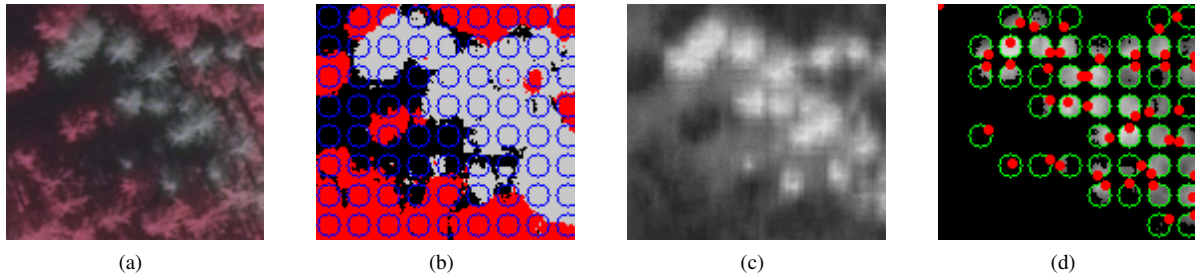


Figure 3. Selecting locations for initializing the level set segmentation. (a) Input image. (b) Result of candidate region detection covered by circles which indicate candidate starting locations. (c) Pixel-level dead tree probabilities of image points from intensity prior. (d) Intersection of candidate circles and dead tree cluster from candidate region detection. Red dots mark chosen starting locations for segmentation, grayscale values depict dead tree probability as in (c).

gradient:

$$\begin{aligned} \frac{\partial \alpha}{\partial t} &= - \int_{\Omega} \omega F(x) \delta[\phi(x)] \begin{pmatrix} \psi_1(x') \\ \vdots \\ \psi_c(x') \end{pmatrix} dx + \frac{1}{\sigma^2} \frac{\sum_{i=1}^M (\alpha_i - \alpha) L_i}{\sum_{i=1}^M L_i} \\ \frac{\partial h}{\partial t} &= - \int_{\Omega} \omega F(x) \delta[\phi(x)] \left(\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2} \right)^T dx \\ \frac{\partial \theta}{\partial t} &= - \int_{\Omega} \omega F(x) \delta[\phi(x)] \left(\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2} \right)^T \frac{\partial R_{\theta}}{\partial \theta} dx \end{aligned} \quad (16)$$

In the above, $x' = R_{\theta}x + h$, $x = (x_1, x_2)$, $L_i = \exp(-\frac{|\alpha - \alpha_i|^2}{2\sigma^2})$, and $\frac{\partial R_{\theta}}{\partial \theta}$ is the component-wise derivative of the rotation matrix R_{θ} by the angle θ (see Tsai et al. (2003) for details). The term $\delta(z)$ denotes the Dirac delta function. In practice, we use a smooth approximation of δ . We apply the standard gradient descent method to update the parameters α, h, θ with a small time step dt until convergence is reached.

2.4 Iterative segmentation

The level set formulation described in Sec. 2.3 allows the segmentation process to converge to at most one object. The location of the final shape is dependent on the initial position of the contour. However, in our scenario we expect to find many dead trees in the input image. Therefore, we develop a procedure to iteratively apply the segmentation at different locations of the image, while ensuring that the individual segmented shapes do not overlap. To achieve this, we first cover the input image with circles C_i of a diameter d approximately equal to that of the containing circle of the largest training polygon, by placing them on a regular grid (see Fig. 3). We then consider each of the circles C_i as a potential location for the initial contour of the level set segmentation. Let C_S be the set of the circles C_i which have a non-empty pixel intersection I_i with the snag cluster obtained in the candidate region detection step (Sec. 2.2). Let $p_i = \operatorname{argmax}_{p \in I_i} \mathcal{P}(l = \text{'dead tree'} | I(x))$ be the pixel in I_i which attains the highest dead tree class probability from the intensity prior (Sec. 2.3.4). We rank the circles in $C_i \in C_S$ according to the probability of their best point p_i and process them in the resulting order. For a circle C_i , we run the level-set segmentation starting with the mean shape $\bar{\phi}$ centered at point p_i . If the segmentation converges to an empty contour, we conclude that there is not enough evidence in the image region around p_i to support the existence of a dead tree. Otherwise, we add the polygon mask obtained from the converged shape s_i to the result set. Note that the image may contain strong local attractors to which the segmentation would converge from starting points within a large radius. To mitigate this, we set the image intensity values of all pixels belonging to s_i to a value which attains a near-zero

probability in our prior intensity distribution. This prevents the formation of overlapping polygons in the segmentation process.

3. EXPERIMENTS

3.1 Material

We tested the proposed approach on three sample plots located in the Bavarian Forest National Park (49°3'19'' N, 13°12'9'' E), which is situated in South-Eastern Germany along the border to the Czech Republic. The study was performed in the mountain mixed forests zone consisting mostly of Norway spruce (*Picea abies*) and European beech (*Fagus sylvatica*). The dead wood originated from an outbreak of the spruce bark beetle (*Ips typographus*) in recent years (Lausch et al., 2013). Color infrared images were acquired in the leaf-on state during a flight campaign carried out in August 2012 using a DMC camera. The mean above-ground flight height was 1900 m, resulting in a pixel resolution of 20 cm on the ground. The images contain 3 spectral bands: near infrared, red and green. The test plots 1,2, and 3 have an area of 618 m², 2426 m², and 7153 m², respectively. These plots were chosen to contain diverse scenarios such as shadows, open ground, clusters of nearby dead trees as well as isolated snags surrounded by living vegetation. The color infrared image of Plot 3 is shown in Fig. 4.

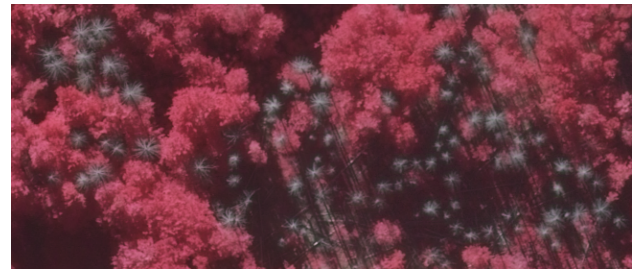


Figure 4. Color infrared image of Plot 3. Shades of red indicate European beech trees, gray hues show standing dead Norway spruce trees.

3.2 Reference data

The reference data for validating the results was obtained by manually marking the contours of the individual dead trees on the three test images. The contours were subsequently converted to polygon masks in order to enable a pixel-based comparison (see Fig. 5). The number of delineated snags was 18, 23 and 83 respectively for Plots 1,2, and 3.

3.3 Evaluation strategy

We utilize the polygon mask representation (Fig. 5(b)) to perform a precise comparison of the detected and reference objects at the

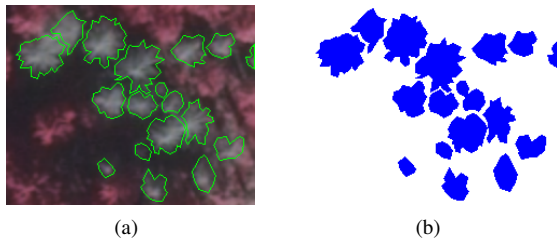


Figure 5. Reference data created using manual delineation in image. (a) Contours of marked polygons. (b) Polygon masks created from contours.

pixel level. We consider a reference snag o_R and a detected snag o_D as potentially matched if their polygon masks have a non-zero pixel intersection. Note that due to over- or under-segmentation, there may generally exist a many-to-many relationship between detected and reference objects. To better describe the quality of a matching (o_R, o_D) , we augment it with the values of two metrics. The first metric is the Euclidean distance d_{ctr} between the centers of gravity of the reference and detected polygons, which are calculated as the average planimetric coordinates of all pixels belonging to the respective polygon. The second metric is the Dice coefficient DSC defined on two sets of pixels A, B as: $DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}$. The Dice coefficient is normalized on the interval $[0; 1]$ and measures the similarity of two sets, with a value of 0 indicating no overlap and a value of 1 indicating set equality. We use two methods to evaluate the quality of the entire detection result: a strict 1-1 correspondence strategy and a lenient N-M correspondence strategy. In the first case, we seek a 1-1 matching between reference and detected trees. We set up a bipartite graph $G = ((V_R, V_D), E)$ with vertices in V_R and V_D corresponding to reference and detected snags, respectively. Each edge $e \in E$ connects a reference and detected vertex. The edge weight, which can be regarded as the assignment cost, is set to the centroid distance between the corresponding polygons if the polygons intersect, and $+\infty$ otherwise. We find the minimum cost assignment using the Kuhn-Munkres algorithm. The 1-1 strategy incorporates information about under/over-segmentation as well as reference snags which were completely missed and detected snags which did not overlap with any reference polygon. The second strategy allows an N-M relationship between reference and detected snags. It considers all potentially matched object pairs (in the sense of pixel overlap) as matched. This strategy is only informative in terms of entirely missed reference and detected dead trees and does not quantify over/under-segmentation issues. However, it is still meaningful to apply it in our scenario, because both the reference and detected polygons are disjoint by construction. For both strategies, we calculate the correctness and completeness measures on the resulting assignment, which are defined as the ratio of the number of matched pairs to the number of, respectively, detected and reference polygons. Finally, we utilize the average per-pixel correctness and completeness. For the pair (o_R, o_D) these measures are defined, respectively, as the ratio of the pixel intersection size between o_R, o_D to the number of pixels in the polygon of o_D and of o_R .

3.4 Experimental setup

As the basis for our experiment, we marked a total of 96 training polygons corresponding to dead trees in images distinct from the test plots. We selected polygons representing diverse lighting conditions and camera perspectives to obtain a reasonable coverage of the target objects' appearance. We then aligned the set of training polygons and derived the eigenshape representation (see Sec. 2.3.3). We used $c = 20$ eigenmodes of shape variation, which represented a total of 98.4% of the variance of the original

shape variability matrix S . We also marked several regions containing only shadows and only living vegetation in the training images, with a total of ca. 20000 pixels. The mean pixel intensities of shadow, living vegetation and snag regions were used as the mean priors in the candidate region detection step (Sec. 2.2). We re-used the labeled dead tree and background polygons for training a kernelized logistic model for use as the intensity prior (Sec. 2.3.4). A total of $T = 3000$ pixel values were utilized as the KLR training points y_i , with the kernel bandwidth selected using 10-fold cross-validation. The overall accuracy of the KLR model obtained from cross-validation was 0.81.

4. RESULTS AND DISCUSSION

Table 1 summarizes the detection performance of our method on the three test plots. The reference and detected contours are outlined in Fig. 6. In Plot 1, most trees are clustered in the central area and the main challenge consists of separating the individual snags. For this plot, all reference trees can be matched to detected snags, with a false positive rate of 0.25. The algorithm does a reasonable job of delineating the individual crowns, attaining an average Dice coefficient value of 0.57 and center-of-gravity displacement of 5.5 pixels (see Table 2). The significant difference between the correctness in the 1-1 and N-M matching strategies reveals an over-segmentation. The false positive polygons mostly lie in dark areas at the perimeter of larger trees. However, note that the dead tree pixel probability within these polygons is relatively high (cf. Fig. 3(c)). The segmentation energy functional only interacts with the pixel intensities through the prior distribution function, and therefore any inaccuracies in the discriminative intensity model, in our case kernelized logistic regression, will be directly transferred to the segmentation result. As stated in Sec. 3.4, the model's pixel classification accuracy was 0.81, which indicates that the snag and background pixels cannot be clearly separated using the supplied 3 spectral channels, perhaps due to diverse lighting conditions and varying hues of the target objects. Recall that in our formulation, we may introduce any generative or discriminative intensity model. We hypothesize that this part of our method could be improved upon by exploring a richer feature space, e.g. augmented by further spectral bands or by transformations thereof such as the NDVI, or by applying more powerful learning methods such as probabilistic graphical models. Plot 2 contains a similar number of snags as Plot 1, however they are distributed across a larger area. This plot tests the ability to detect isolated snags surrounded by living vegetation or standing in shadowed areas. The algorithm is able to delineate the individual snags quite reliably, achieving a completeness of 0.96 with a correctness of 0.71. Part of the false positives detected in this plot are associated with small open ground patches visible in the image that additionally contain fallen trees. Once again this deficiency may be attributed to the intensity classifier, because we did not include ground or fallen tree pixels in the training set, and their colors in the image are closer to the snag class than to the remaining background classes. These problems outline a disadvantage of our purely image-based approach versus methods which have access to a surface model or a 3D point cloud, where such open ground patches could be easily detected using simple thresholding schemes. As expected, the quality of the detected polygons is higher in this plot compared to Plot 1, because finding the boundary of single crowns is an easier and less ambiguous task than splitting a cluster of several crowns. Therefore, the mean Dice coefficient is 12 percentage points higher at 0.69 and the mean center-of-gravity displacement more than a pixel lower at 4.4 px. Plot 3, our final test area, is quite challenging due to the high variability of the snag crown diameters (Fig. 4), presence of shadowed areas, open ground patches and fallen dead trees. Nev-

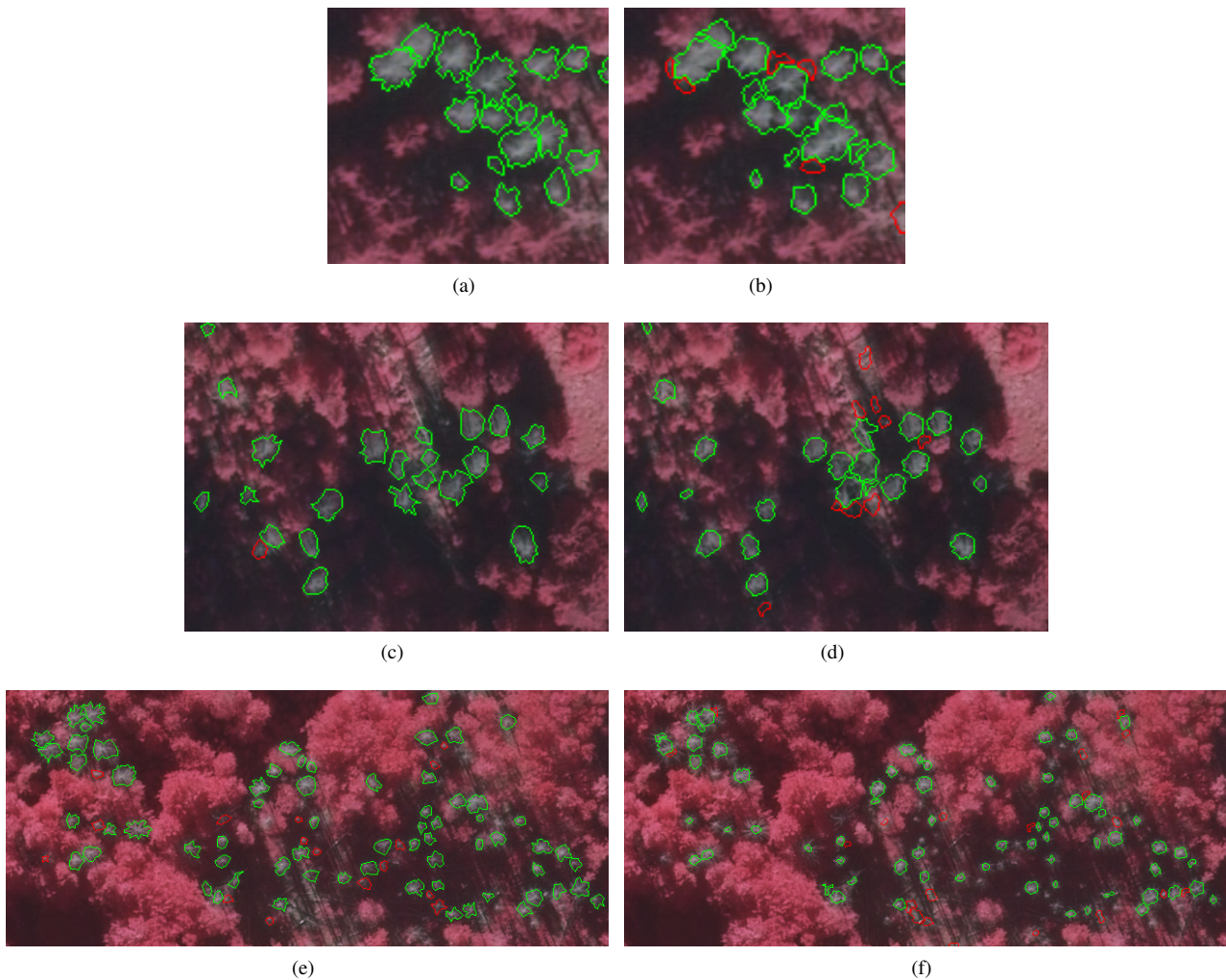


Figure 6. Reference and detected polygons. (a), (c), (e): reference polygons respectively of Plot 1, 2, and 3. (b), (d), (f): detected polygons respectively in Plot 1, 2, and 3. Red contours denote unmatched, green denote matched polygons using the 1-1 matching strategy. Note the different image scales - (a),(b): 26 m x 24 m; (c),(d): 58 m x 42 m; (e),(f): 130 m x 55 m.

ertheless, our algorithm is able to attain a completeness value of 0.81 with a correctness of 0.77. A drop in pixel-level completeness to 0.5 (see Table 2) can be observed in comparison to the other plots. This can be partially explained by the presence of snags with elaborate branch structures visible in the image. The pixel classifier does not take any neighborhood properties of a pixel into account, and therefore any fine details are usually lost when calculating the snag probabilities (Fig. 7). This makes it impossible for our method to recover the fine branch structure, and instead only the core area of the crown is delineated.

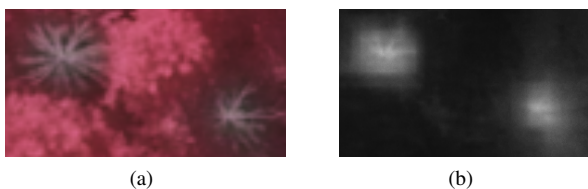


Figure 7. Loss of detail due to transformation into probability image.

Since it is unlikely that a perfect point-level snag classifier can be obtained, it seems more promising to alter the segmentation energy functional to include a term which evolves the intensity model together with the shape using contrast information present in the original image (such as the classical Chan-Vese model,

Chan and Vese (2001)) while maintaining a balance with the intensity prior. This is expected to come at an additional computational cost.

Metric	Plot 1	Plot 2	Plot 3
Correctness (1-1)	0.75	0.71	0.77
Completeness (1-1)	1	0.96	0.81
Correctness (N-M)	0.96	0.77	0.85
Completeness (N-M)	1	1	0.82

Table 1. Evaluation of detection results - correctness and completeness for object-level detection.

5. CONCLUSIONS

An integrated method for detecting standing dead trees from aerial imagery with near infrared channels was presented. We proposed a two-stage approach where the approximate snag location regions are first segmented using a Gaussian Mixture Model (GMM) with priors on the region intensity means, following which the candidate regions are refined by an iterative application of a level set segmentation algorithm, yielding a set of delineated polygons corresponding to the individual found snags. The level set segmentation procedure incorporates shape and intensity prior

Metric	Plot 1	Plot 2	Plot3
Dice Coefficient	0.57	0.69	0.62
Center-of-gravity distance [px]	5.5	4.4	3.5
Per-pixel correctness	0.59	0.76	0.89
Per-pixel completeness	0.61	0.67	0.50
Per-pixel correctness (N-M)	0.76	0.58	0.78
Per-pixel completeness (N-M)	0.85	0.62	0.52

Table 2. Evaluation of detection results. Rows 1-4: average polygon matching quality metrics for 1-1 matching. Rows 5-6: average per-pixel correctness and completeness for N-M matching.

information which was obtained from manually labeled training data. We tested our method on three plots with reference data created manually by visual inspection. Additionally to the standard detection rate metrics, the structure of our reference and result objects (polygon masks) allowed for a precise, pixel-wise evaluation of the similarity and displacement between object centers. Our results indicate that the proposed algorithm is able to achieve a correctness of 0.71-0.77 and completeness of 0.8-1.0 for the snag detection task, with a center displacement of 3-5 pixels and an average Dice coefficient between 0.57 and 0.69. Two potential problem areas are the difficulty in distinguishing between open ground, lying dead trees and snags purely based on color information, as well as the loss of fine detail because of the transformation of the image intensity values into probability values. The former problem could be alleviated by the introduction of an additional 3D elevation channel which would facilitate filtering out ground patches and fallen stems, and also by enriching the set of features for point-based classification. As for the latter issue, it could be beneficial to incorporate a data-driven intensity term into the energy functional to make better use of contrast regions within the input image. Resolving these two problems and carefully choosing training examples which capture the full extent of the dead trees' variability should make it possible to further improve upon the completeness and correctness rates. Finally, the cluster information discovered by the GMM in the first step could be re-used to a greater extent by the level set segmentation, instead of merely being a way to filter out unpromising regions of the image to improve computational performance.

REFERENCES

- Bhattacharai, N., Quackenbush, L. J., Calandra, L., Im, J. and Teale, S. A., 2012. An automated object-based approach to detect Sirex-infestation in pines. In: *ASPRS Annual Conference Proceedings*.
- Bütler, R. and Schlaepfer, R., 2004. Spruce snag quantification by coupling colour infrared aerial photos and a GIS. *Forest Ecol. Manag.* 195(3), pp. 325 – 339.
- Chan, T. and Vese, L., 2001. Active contours without edges. *IEEE T. Image Process.* 10(2), pp. 266–277.
- Cremers, D. and Rousson, M., 2007. Efficient kernel density estimation of shape and intensity priors for level set segmentation. In: *Deformable Models, Topics in Biomedical Engineering*. International Book Series, Springer, New York, pp. 447–460.
- Cremers, D., Osher, S. and Soatto, S., 2006. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. Comput. Vision* 69(3), pp. 335–351.
- Cremers, D., Rousson, M. and Deriche, R., 2007. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *Int. J. Comput. Vision* 72(2), pp. 195–215.
- Eskelson, B., Temesgen, H. and Hagar, J., 2012. A comparison of selected parametric and imputation methods for estimating snag density and snag quality attributes. *Forest Ecol. Manag.* 272(0), pp. 26 – 34.
- Fraley, C. and Raftery, A. E., 2007. Bayesian regularization for normal mixture estimation and model-based clustering. *J. Classif.* 24(2), pp. 155–181.
- Fridman, J. and Walheim, M., 2000. Amount, structure, and dynamics of dead wood on managed forestland in Sweden. *Forest Ecol. Manag.* 131(1-3), pp. 23–36.
- Harmon, M., Franklin, J., Swanson, F., Sollins, P., Gregory, S., Lattin, J., Anderson, N., Cline, S., Aumen, N., Sedell, J., Lienkaemper, G., Jr., K. C. and Cummins, K., 1986. Ecology of coarse woody debris in temperate ecosystems. *Advances in Ecological Research*, Vol. 15, Academic Press, pp. 133 – 302.
- Heurich, M., Ochs, T., Andresen, T. and Schneider, T., 2010. Object-orientated image analysis for the semi-automatic detection of dead trees following a spruce bark beetle (*Ips typographus*) outbreak. *European Journal of Forest Research* 129(3), pp. 313–324.
- Jensen, J. R., 2006. *Remote Sensing of the Environment: An Earth Resource Perspective* (2nd Edition). Prentice Hall, Upper Saddle River.
- Ke, Y., Zhang, W. and Quackenbush, L. J., 2010. Active contour and hill climbing for tree crown detection and delineation. *Photogramm. Eng. Rem. S.* 76(10), pp. 1169–1181.
- Kim, Y., Yang, Z., Cohen, W. B., Pflugmacher, D., Lauver, C. L. and Vankat, J. L., 2009. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sens. Environ.* 113(11), pp. 2499 – 2510.
- Lausch, A., Heurich, M. and Fahse, L., 2013. Spatio-temporal infestation patterns of *Ips typographus* (L.) in the Bavarian Forest National Park, Germany. *Ecological Indicators* 31, pp. 73–81.
- Lee, S. and Cho, H., 2006. Detection of the pine trees damaged by pine wilt disease using high spatial remote sensing data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36, Part 7.
- Leventon, M., Grimson, W. and Faugeras, O., 2000. Statistical shape influence in geodesic active contours. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 316–323.
- Pasher, J. and King, D. J., 2009. Mapping dead wood distribution in a temperate hardwood forest using high resolution airborne imagery. *Forest Ecol. Manag.* 258(7), pp. 1536 – 1548.
- Tsai, A., Yezzi, A., III, W. W., Tempany, C., Tucker, D., Fan, A., Grimson, W. and Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE T. Med. Imaging* 22(2), pp. 137–154.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, pp. 127–150.
- Vogelmann, J., 1990. Comparison between two vegetation indices for measuring different types of forest damage in the north-eastern united states. *Int. J. Remote Sens.* 11(12), pp. 2281–2297.
- Wang, C., Lu, Z. and Haithcoat, T. L., 2007. Using Landsat images to detect oak decline in the Mark Twain National Forest, Ozark Highlands. *Forest Ecol. Manag.* 240(13), pp. 70 – 78.
- Woodall, C., Heath, L. and Smith, J., 2008. National inventories of down and dead woody material forest carbon stocks in the United States: Challenges and opportunities. *Forest Ecol. Manag.* 256(3), pp. 221–228.
- Yao, W., Krzystek, P. and Heurich, M., 2012. Identifying Standing Dead Trees in Forest Areas Based on 3D Single Tree Detection From Full Waveform Lidar Data. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 1-7, pp. 359–364.
- Ye, J., Wittman, T., Bresson, X. and Osher, S., 2010. Segmentation for hyperspectral images with priors. In: *Advances in Visual Computing*, Lecture Notes in Computer Science, Vol. 6454, Springer, Berlin, pp. 97–106.