Machine Vision and Applications (2007) 18:317–330
DOI 10.1007/s00138-006-0064-9

ORIGINAL PAPER

Automatic extraction and delineation of single trees from remote sensing data

Bernd-Michael Wolf (né Straub) · Christian Heipke

Received: 19 September 2005 / Accepted: 14 November 2006 / Published online: 24 January 2007

Abstract In this paper, we present a novel approach for the automatic extraction of trees and the delineation of the tree crowns from remote sensing data, and report and evaluate the results obtained with different test data sets. The approach is scale-invariant and is based on co-registered colour-infrared aerial imagery and a digital surface model (DSM). Our primary assumption is that the coarse structure of the crown, if represented at the appropriate level in scale-space, can be approximated with the help of an ellipsoid. The fine structure of the crown is suppressed at this scale level and can be ignored. Our approach is based on a tree model with three geometric parameters (size, circularity and convexity of the tree crown) and one radiometric parameter for the tree vitality. The processing strategy comprises three steps. First, we segment a wide range of scale levels of a pre-processed version of the DSM. In the second step, we select the best hypothesis for a crown from the overlapping segments of all levels based on the tree model. The selection is achieved with the help of fuzzy functions for the tree model parameters. Finally, the crown boundary is refined using active contour models (snakes). The approach was tested with four data sets from different sensors and exhibiting different resolutions. The results are very promising and prove the feasibility of the new approach for automatic tree extraction from remote sensing data.

Keywords Tree extraction · Scale space · Automatic · Snakes · Real imagery

1 Introduction

Trees are important topographic objects in different fields of applications. Obviously, trees play an essential role in forestry, and forest inventories regularly capture tree data all over the world. In addition, ecological and social aspects are increasingly important: forests provide shelter for many different species of the Earth's flora and fauna. They are a major reservoir for CO_2 storage and green areas including parks and single trees are considered important factors for making a city attractive. Information about the position and size of trees in cities is also beginning to be used in computer graphics aimed at producing more realistic views of the environment.

The first trial of the use of aerial images for forest purposes was performed in 1897 according to Hildebrandt [11]. Since that time the forest scientific community has been working on improving methods for the extraction of tree parameters from aerial images such as the tree type and health status, the height and crown size, the stem diameter at breast height, and the stem volume [12]. Early work concentrated on the manual interpretation of images for forest inventory [17,24]. Since then a number of attempts to automate this task have appeared in the literature (see Sect. 2).

In this paper, we present a new approach for the automatic extraction of individual trees using a

The major part of this work was carried out while both authors worked together at the Institute of Photogrammetry and GeoInformation, University of Hannover.

B.-M. Wolf (né Straub)
SOLVing3D GmbH, Osteriede 5, 30927 Garbsen, Germany

C. Heipke (✉)
Institute of Photogrammetry and GeoInformation,
University of Hannover, Nienburger Str. 1,
30167 Hannover, Germany
e-mail: heipke@ipi.uni-hannover.de

geo-referenced colour-infrared aerial image[1] and a digital surface model (DSM)[2] as input data (see also [26]). The approach is based on a geometric and radiometric model of a tree, a segmentation at multiple scales followed by scale selection and a refinement step using active contour models (snakes). The mathematical reasoning is mainly based on differential geometry. The main data source in our approach is the DSM. Additional colour information from the image is used to differentiate between vegetation and other objects in the scene. Our aim is to detect every tree in the observed part of the real world and to determine the boundary of their crowns.

This paper is organised as follows: in the next section an overview is given of related work in automatic tree extraction from remote sensing data in forest and settlement areas. In the main section of the paper we describe our approach. The description is divided into two subsections: in the first subsection the object model for trees is presented, and in the second the processing strategy is elaborated. In the last section, we show results from four different data sets including a quantitative evaluation based on independently obtained reference data. We close with a short summary and an outlook.

2 Related work

The first approaches to automatic extraction of individual trees from images were proposed by pioneers in the field almost two decades ago [9,10,20]. More recent work in the field is reported in Gougeon [8], Pollock [20,21], Brandtberg and Walter [4], Andersen et al. [1], and Persson et al. [19]. A good overview of the field is contained in [13]. Some of the important publications are described in detail in this section.

A common element of most approaches is the geometric model of a tree, which was proposed by Pollock [21,22]. Pollock describes the tree geometry by means of a generalised ellipsoid of revolution, see Eq. (1). In the following, this model is termed *Pollock-Model*,[3] and the corresponding synthetic trees are termed *Pollock-Trees*.

$$\frac{z^n}{a^n} + \frac{\left(x^2 + y^2\right)^{\frac{n}{2}}}{b^n} = 1 \tag{1}$$

In Eq. (1) (x, y, z) are point co-ordinates on the crown surface. a corresponds to the tree height, and b to the crown radius. n is a shape parameter. According to Gong et al. [7], a typically ranges from 15 to 40 m, and b from 2.5 to 15 m. Form parameters n for deciduous trees lie in the range of 1.0–1.8 with a typical value of 1.2, and for coniferous trees in the range of 1.5–2.5 with a typical value of 2.0. Two different surfaces, which can be described with Eq. (1), are depicted in Fig. 1: the left one is an example for a deciduous tree, and the right for a coniferous tree.

The surface of a real tree is of course more complex than the one described by the Pollock-Model, because a tree may not exhibit circular growth due to neighbouring obstacles or wind. Individual branches have different length and a tree crown may be split into smaller sub-crowns. This fine structure is not captured in the Pollock-Model. Nevertheless, the main shape of the upper part of the crown is reasonably well represented. The Pollock-Model has been the basis of a number of algorithms developed in recent years, partly in a refined form. Examples include Larsen and Rudemo [16], Straub and Heipke [27], Andersen et al. [1], Gong et al. [7], Weinacker et al. [29], and Wulder et al. [31].

In general, there are two possibilities for building a strategy for the automatic extraction of trees from image or surface data. The first possibility is to model the crown in detail: one could try to detect and group the fine structure in order to reconstruct the individual crowns. The second possibility is to remove the fine structures from the data with the aim of creating a surface, which has the character of the Pollock-Model. Both strategies can be found in the literature: Brandtberg [3] proposed the use of the typically fine structure of deciduous trees in optical images for the detection of individual trees. The other strategy, removal of the fine structure, was proposed in Schardt et al. [23] and in Persson et al. [19]. In Andersen et al. [1] the fine structure of the crown is modelled as a stochastic process with the aim of detecting the underlying coarse structure of the crown. The main problem of this type of approach is the determination of an optimal low pass filter for every single tree in the image. This is analogous to a "chicken-and-egg" problem, because the optimal low pass filter depends mainly on the diameter of the individual tree being sought, which is not known in advance. Brandtberg and Walter [4] proposed a solution to this problem by using a linear scale space representation of the image.

The basic idea of linear scale space is to construct a multi scale representation of an image, which only depends on one parameter and has the property of *causality*: features at a coarse scale must have a cause at fine scale [15]. The scale space transformation itself should

[1] In a georeferenced aerial image for each pixel world coordinates are available in planimetry. Such an image is also called an "orthoimage" or "orthophoto".

[2] We use a gridded version which was obtained either by image matching or by aerial laser scanning.

[3] Note that the Pollock-Model is introduced here due to its significance in the field. Although we do not explicitly use this model in our approach, our ideas have been influenced by the work of Pollock, see also Sect. 3.

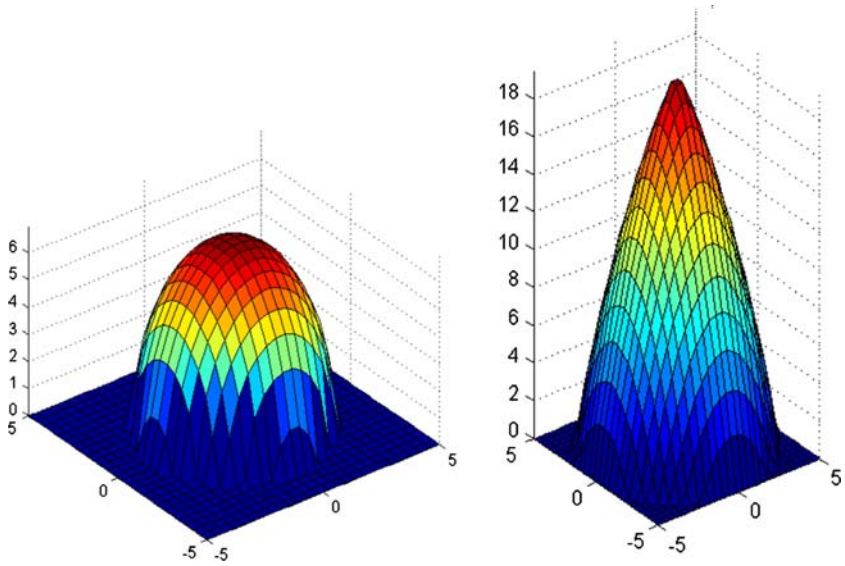

Fig. 1 3D visualisation of the Pollock-Model. *Left* surface model of a typical deciduous tree: $a = 7, b = 3.5, n = 2.0$ *right* coniferous tree $a = 20.0; b = 5.0; n = 1.2$

not lead to new features. One can show that a multi-scale representation based on a Gaussian function as low pass filter fulfils this requirement.[4] In practice, the original signal is convolved with a Gaussian with different scale parameters σ. Small values of σ correspond to a fine scale, large values to a coarse scale. An extensive investigation on linear scale space transformations of images, mathematical reasoning and technical instructions can be found in [18].

Quantitative results have been reported for some of the discussed approaches. In Persson et al. [19] 71% of the trees were extracted, and all extracted objects were trees (71% completeness, 100% correctness). Pollock [22] achieved a completeness of 61% and a correctness of 85%. Better results are reported in Brandtberg and Walter [4] with a completeness of 85% and a correctness of 100%, and in Andersen et al. [1] with a completeness of 83% and a correctness of 89%. While these percentages give a first indication of the quality of the approach, and some of them are rather impressive, it should be noted that they are hard to compare, since they relate to different test sites with different scene complexity, different input data, different image resolutions, different numbers of trees per data set and different evaluation strategies.

3 Description of the approach

A critical point in object extraction from images is the selection of the appropriate scale level. The reasons are: (1) the correct scale level mainly depends on the size of the object one is looking for. The size of trees depends on the age, the habitat, the species and many more parameters, which in general cannot be modelled in advance. Therefore, the tree size and thus the correct scale level can neither be assumed to be known a priori nor is it constant for all trees in one image. (2) The correct scale is of crucial importance for image segmentation. The fine structure of a tree crown is very difficult to model but—apart from this fine structure—the crown has a relatively simple shape.

In our approach the image is segmented in a wide range of scales across the linear scale space, just bounded by reasonable values for the minimum and the maximum radius of a tree crown (see also Sect. 2). The operator we use for segmentation is essentially a Laplacian-of-Gaussian and thus the bandwidth of this operator is 1.2 octaves. We increase the scale parameter in steps starting from $2^{0.5i}$ m with $i = 1, 2, 3, 4, 5, 6$, and not in a power-of-two series, which may seem to be needlessly inefficient. However, we later select the best tree hypothesis based on overlapping segments (see below), which requires redundant information between the different levels.

In Sect. 3.1 a detailed description of the model of individual trees is given. In Sect. 3.2 the processing strategy for the extraction of these trees from the image and height data is described.

3.1 Tree model

3.1.1 Geometric properties

The coarse structure of the crown is implicitly modelled using the Pollock-Model. For all possible shapes, which can be represented with this model, the projec-

[4] Strictly speaking this is only true in 1D. In 2D there are exceptions such as the "dumbbell image", which, however, do not have a noticeable impact for our application.

tion into the xy-plane is a circle with a diameter in a pre-defined range, furthermore the 3D shape of the surface is convex.

Convexity is measured as the mean value of the Laplacian inside a segment. The justification for this is explained as follows. We start with investigating the DSM of four synthetic Pollock-Trees, denoted $H(x, y)$. In the left part of Fig. 2 these four Pollock-Trees computed with $a = 6$ (m), $b = 2$ (m), and $n = 2.0$ are depicted. As is often the case the trees stand relatively close together and increasingly so from top to bottom. In the right part of Fig. 2 various profiles are shown, which were taken along the white line visible in the left part of the figure. The leftmost profile point is the uppermost point of the white line.

The profile of the DSM is plotted in dark grey. One can see that the value of the local minima between the trees increases from left to right, and the gradient magnitude (black line in Fig. 2) decreases. Obviously, this is a consequence of the decreasing distance between the trees, and of the crown's shape. The surface at the treetops has a convex shape in both directions, along and across the profile. Therefore the Laplacian (see Eq. 2), which is depicted in light grey in Fig. 2, is negative for the whole crown.

$$\frac{\partial^2 H(x,y)}{\partial x^2} < 0 \quad \wedge \quad \frac{\partial^2 H(x,y)}{\partial y^2} < 0 \quad \Rightarrow \quad \frac{\partial^2 H(x,y)}{\partial x^2} + \frac{\partial^2 H(x,y)}{\partial y^2} < 0 \tag{2}$$

At the points on the profile between two trees the second partial derivative is less than zero along the profile and greater than zero perpendicular to the profile. Therefore, the Laplacian is generally larger than at points within the crown. These characteristics lead to local maxima of the Laplacian between the crowns.

In the case of real data this model is only valid in some scale levels. A height profile from real data is used to explain this statement. Two different scale space representations of the surface model $H(x, y)$ are depicted in Fig. 3. One can see that more and more of the fine structures disappear and the coarse structure is revealed when the scale parameterσ is increased.

The height profile along the treetops is determined along the dotted line, which is superimposed on the surface model in Fig. 3. The left height profile, which was taken from the original surface model, contains more high frequency information compared to the profile of the synthetic trees. As a result the Laplacian oscillates close to zero. At the coarser scale level the mean convexity can be used as geometric feature for the decision, if a segment is a crown or not. Similar to the profile of the synthetic Pollock-Trees the Laplacian is negative for tree crowns, and borders between the trees represent local maxima.

3.1.2 Radiometric properties

Vegetation exhibits a typical response in the red and the near infrared band of the electromagnetic spectrum. The reflection of the near infrared band is higher for vegetation than for areas without vegetation and in the red band it is lower. Vegetation indices make use of this property. We use the *normalized difference vegetation index* (NDVI, see for example [17]) to differentiate vegetation and non-vegetation areas in the images (Eq. (3). *NIR* is the grey value in the near infrared band, and *red* the grey value in the red band. Refer to Fig. 4 for an example.

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{red}}{\mathrm{NIR} + \mathrm{red}} \tag{3}$$

Various improvements for computing the NDVI were suggested over the years. A particularly interesting one was described by Ünsalan and Boyer [28], in which the authors address the effect of saturation by linearising the vegetation index based on a statistical framework. Since we combine the NDVI with a membership function in the evaluation step (see below), we can in principle achieve the same effect by properly defining the membership function.

3.2 Processing strategy

Our processing strategy comprises three steps: first, we segment a wide range of scale levels of a pre-processed version of the DSM. The segmentation is achieved by applying the watershed transformation [2,25]. In a subsequent step, we evaluate the segments and select the best hypothesis for a crown in case various segments from different scale levels overlap. The evaluation and the selection of the best hypotheses are achieved using fuzzy functions [32] for the tree model parameters. Finally, the crown boundary is refined using snakes [14]. In the following, we describe how to combine these tools with the aim of detecting individual trees and reconstructing the outline of the crown.

3.2.1 Segmentation of the surface model

As mentioned above, we perform a segmentation of the surface model in a wide range of scales. The segmentation itself should be free of parameters and work only in the image space and not in the feature space, because the feature space is not independent of the scale level. The

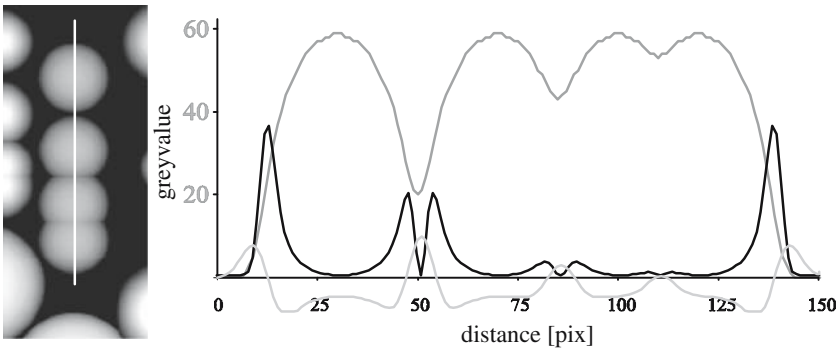

Fig. 2 Height values (*left*) and various profiles (*right*) of the surface model of four Pollock-Trees. The *dark grey line* represents the height values, the *black line* shows the gradient magnitude, and the *light grey line* depicts the values of the Laplacian

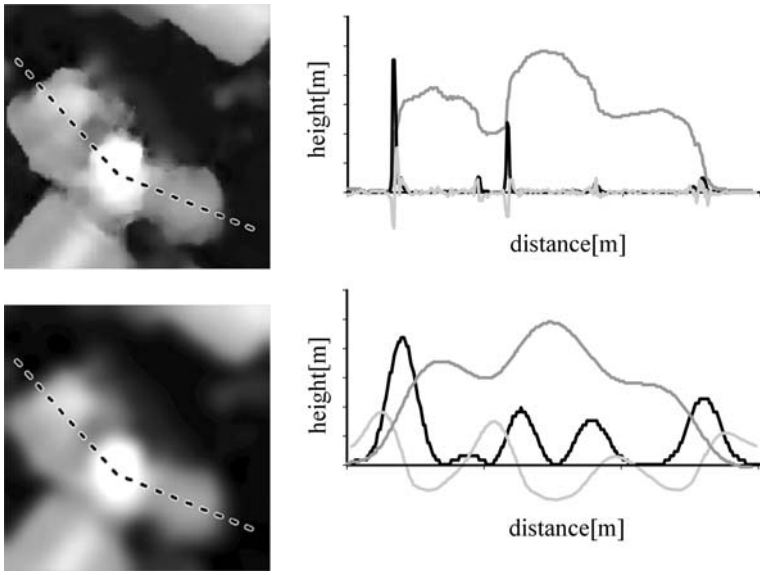

Fig. 3 Representation of the surface model $H(x, y)$ at two different scale levels (*left*). The height profiles to the right were determined along the *dotted line* in the images. *Dark grey* height profile; *black* gradient magnitude, *light grey* Laplacian

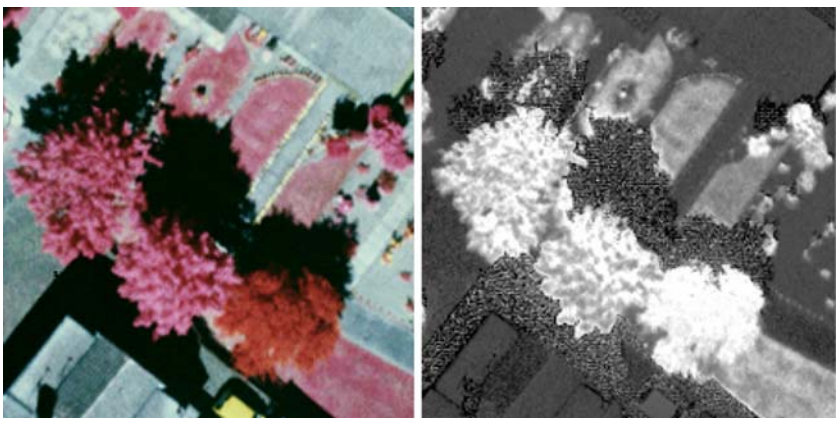

Fig. 4 Example of a colour-infrared image of trees (*left*) and the NDVI of the same area (*bright* areas correspond to positive NDVI values)

watershed transformation fulfils these requirements. In addition it is well suited for the segmentation of height data. One reason is that the key idea of the watershed transformation is the segmentation by means of a flooding simulation [25]. *Basins* are the domains of the image, which fill up first if a "water level" increases starting from the lowest grey value in the image. *Watersheds* are embankments between the basins. It is known that the watershed transformation has a tendency for over-segmenting the images, but since we use a variety of different scales, i.e. differently smoothed DSMs, this fact is compensated for through the choice of the degree of smoothing.

If the watershed transformation is to be applied to extract trees from height data the surface model has to be pre-processed in such a way that the trees themselves

form basins. The easiest way to do this is to invert the DSM, since trees can be considered as blobs. In forest areas this approach works rather well because trees stand very close to each other. In other areas, however, parts of the ground are also visible. Since the distance between the trees can be larger in non-forest areas, the outlines of the basins derived from the inverted surface model are only poor approximations of the crown boundaries (Fig. 5, left).

Better results are achieved, if derivatives of the image function are used (see also Figs. 2, 3). In a number of experiments we have found that first squaring and then inverting the Laplacian constitutes an appropriate preprocessing procedure. While this choice must be characterised as rather heuristic, we have adopted it on the basis of the experimental results.

An example is shown in Fig. 5. The watersheds of the inverted surface model are superimposed on the surface model in the left part of the figure, and the watersheds of the squared and inverted Laplacian in the right part. Note that the basins in the right part of the figure fit much better to the individual tree crowns than the basins in the left part.

3.2.2 *Evaluation of segments and selection of best hypothesis*

Definition of membership functions In this section, we describe the evaluation of the derived segments and the selection of the best hypothesis in case of overlapping segments. Both procedures are based on the tree model parameters size, circularity, convexity, and vitality. Membership functions for the four parameters transform the parameter values into membership values, upon which the evaluation and the selection are based. As explained below, the selection of the membership functions is governed by general knowledge about trees. In order not to introduce additional interpolation effects, we are using piecewise linear functions.

The segment **size** can be derived from the number of pixels in the segment and the ground sample distance of the surface model. The membership function of the tree crown is depicted in the left part of Fig. 6. The lower border is 20 m^2 corresponding to a crown radius of 2.5 m and the upper border is 700 m^2 corresponding to a radius of 15 m. Both size values result in a membership value of 0.75. For smaller and for larger sizes the membership value decreases linearly. The largest possible radius is assumed to be 35 m corresponding to 3,850 m^2. These values for the crown radius cover all tree species [7]. The **circularity** of a segment is computed according to Eq. (4)

$$\text{circularity} = \text{Size} \Big/ \left(\pi r_{\max}^2\right) \tag{4}$$

$r_{\max}$ is the largest distance between the centre of gravity and the border of the segment. A suitable lower border is close to the value of 0.7, which is the circularity value of a square. According to our experience, typical values lie above 0.85, and the upper border is equal to 1. The resulting membership function is depicted in the right part of Fig. 6.

Convexity is captured by investigating the mean value of the Laplacian of the segment. For trees this value must be negative (see also Sect. 3.1.1). Other objects such as buildings and roads consist of many planar patches, and the Laplacian of the resulting segments is close to zero. A negative mean value of the Laplacian of a segment leads to a membership value of 1 and in the case of a positive mean value the membership value is 0 (see left part of Fig. 7).

The tree **vitality** is derived from the optical image. It is used to discriminate vegetation and non-vegetation areas. As mentioned above we use the NDVI value as indicator for the vitality (see right part of Fig. 7). In general, trees have relatively high NDVI values. Therefore, we use a membership function with increasing membership value for positive NDVI values.

Evaluation and selection of hypothesis The evaluation of the segments is subdivided into two steps. First, valid tree hypotheses are selected according to the membership values of the segments. A tree is an object with a defined size, circularity, convexity and vitality. Consequently, the minimum of the four membership values of a segment is defined as the *final membership value* of a segment. This value is tested against a threshold to decide whether or not we consider a segment to be a tree hypothesis. The value for the threshold was found experimentally. Segments with a final membership value higher then 0.5 are considered tree hypotheses.

In a number of cases the tree hypotheses from different scale levels overlap (see Fig. 8). The left image shows the tree hypotheses at a scale level of $\sigma = 1.9$ m and the right one at $\sigma = 3.7$ m. One can see that many hypotheses occur at more than one scale. In some cases the segments are quite similar in both depicted scale levels. In other cases the hypotheses are subdivided at the finer scale level.

In the second step these different situations for every hypotheses are analysed. In accordance with the work of Winter [30] we postulate that two hypotheses B_σ from different scale levels for which the overlap parameter $ov > 0.5$ (see Eq. 5), are hypotheses for the same tree in the real world.

$$ov = \text{size}\,(B_{\sigma_1} \cap B_{\sigma_2}) \;/\; \text{MIN}(\text{size}(B_{\sigma_1}, B_{\sigma_2})) \tag{5}$$

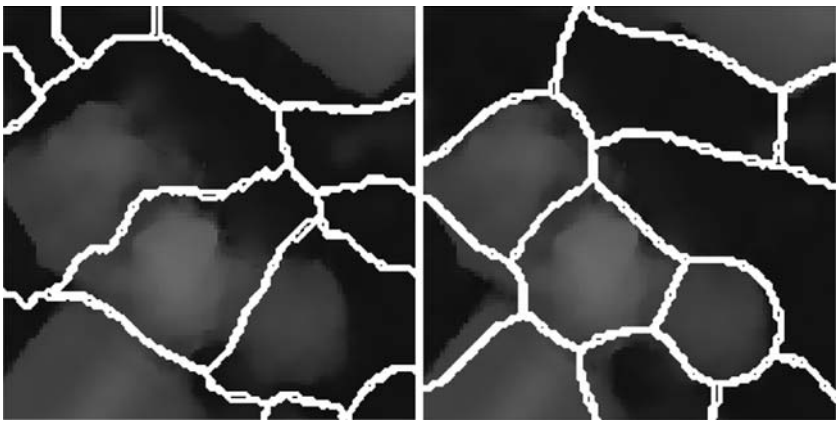

Fig. 5 A subset of the surface model showing three trees with superimposed watersheds resulting from the inverted DSM (*left*) and the squared and inverted Laplacian (*right*)

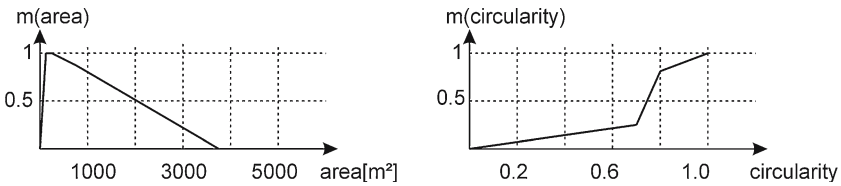

Fig. 6 Membership functions for size (*left*) and circularity (*right*)

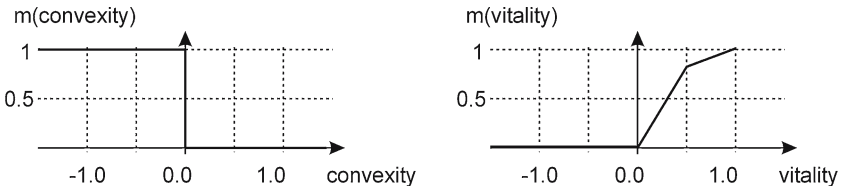

Fig. 7 Membership functions for convexity (*left*) and vitality (*right*)

For overlapping hypotheses the best one is assumed to be the one with the highest final membership value. These finally selected hypotheses are depicted in Fig. 9. The white lines correspond to the outlines of the segments, which are selected as valid trees. Most of the trees in the scene were detected correctly, but many of the boundaries are not very precise approximations for the outline of the individual crowns. This observation leads to the last processing step: the outlines of the crowns are refined using snakes.

3.2.3 Refinement of the tree crown boundary

Whereas the outlines of the segments were extracted at different scale levels, the outline of the crown is an object without a changing scale. We use *snakes* for this task. A snake [14] is a deformable geometric model with physical properties like elasticity. It can also be thought of as a virtual rubber cord, which can be used to detect valleys in a hilly landscape with the help of gravity. If the snake is initialised close to the valley of the landscape, the gravity drags the snake into the valley. The virtual landscape may be represented by a DSM, a grey value image or an edge image. The movement originates from a field of gradients. Such a situation is shown in Fig. 10. In the background one can see the edge of a circular object. The enlargement in the foreground shows the field of gradient vectors.

In general, there are two main drawbacks for the application of snakes as a measurement tool. The first one is that the snake has to be initialised very close to the features one is looking for. Otherwise the behaviour of the snake is nearly impossible to predict. The second one is the tuning of the parameters, primarily the weighting between internal and external energy and the selection of the external energy field itself.

In our approach the snake is only used only for fine measurement in the last stage. The coarse shape of the crown is already known. Furthermore, we have observed that the approximation is often too small. Smaller regions often receive a better evaluation, which seems to be a result of the membership value for the circularity. Based on these constraints we have built a snake which is rather stable under these special conditions: the geometry of the snake is initialised for every tree hypothesis as a closed polygon at the centre of gravity of the corresponding Basin. This initialisation stage is depicted in Fig. 10 as a circle in the right image. The parameters for

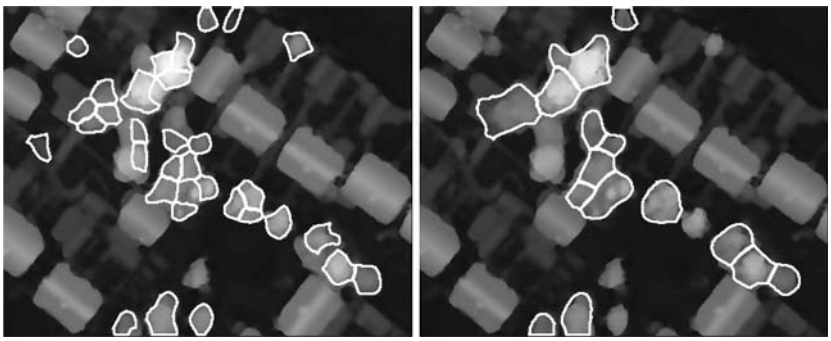

Fig. 8 Valid tree segments at two different scale levels. *Left* σ=1.9 m, *Right* σ=3.7 m

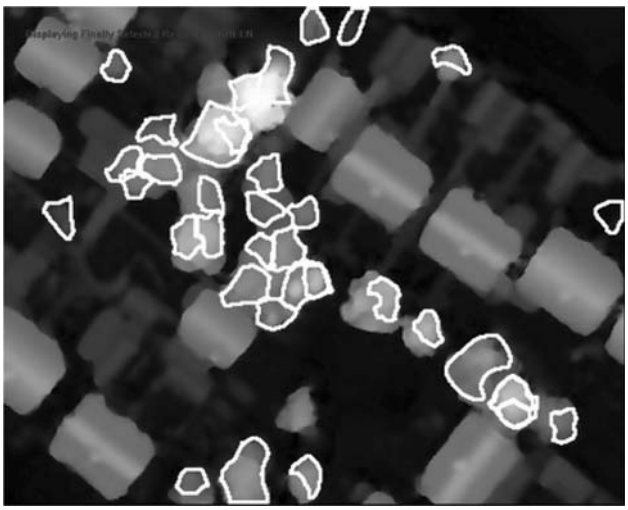

Fig. 9 Finally selected hypotheses superimposed on the surface model. *White lines* depict the best hypotheses over all investigated scales

the internal energies were tuned such that the length of the contour has a low weight and the curvature obtains a high weight. Without external energy a snake which is tuned in such a way converges to a circle with a tendency to shrink. As mentioned, the approximation is often too small. Therefore, an additional energy term is added, which makes the snake behave like a balloon [5]. With this additional term the contour moves towards the outline of the crown even if no external energy influences the movement. The external energy itself is proportional to the gradient magnitude (Fig. 10). In order to enlarge the radius of convergence we smooth the gradient image with a Gaussian kernel before computing the external energy. For illustration purposes the external energy before smoothing is depicted in the left part of Fig. 11 and the one we use in the right part of Fig. 11. The smoothing effect resulting in an enlarged radius of convergence is clearly visible.

Finally, the final membership values of every tree were recomputed because its outline has changed. Also, the overlap computations were repeated. Based on the updated membership values, some hypotheses were rejected because they no longer met the minimum requirements for a tree.

4 Performance evaluation

4.1 Test data and evaluation procedure

In order to investigate the potential of the approach we applied it to four different test data sets. The first one is a high resolution data set depicting the village of *Grangemouth* in Scotland. The images were taken using a conventional aerial camera with a colour-infrared film and were converted into digital images in a separate step using a high quality photogrammetric scanner. We had an orthophoto with a ground sampling distance (GSD) of 0.1 m and a DSM with a grid width of 0.2 m available to us. The DSM was derived from available stereo images using an automatic image matching algorithm including manual editing to remove any blunders [6]. The scene is composed of a total of 160 trees, a number of them single trees, some touching each other slightly, and a few standing in a compact group. This data set can be regarded as the optimal information for the task, both in terms of geometric resolution and DSM quality. Therefore, we expect the data set to produce the best results.

The other three data sets were selected to test the performance of the algorithm in somewhat less ideal situations. In two of the three cases, the GSD of the colour-infrared image was 0.5 m, while one data set did not contain brightness images. The DSM grid width was 1 m in all three cases.

The *Hohentauern* test site lies in a forest in the Austrian Alps. 94% of the trees in the test site are spruce. The surface model was derived from first-pulse laser scanning data. For this data set brightness images were not available. However, since the scene only contained trees (a total of 87), a separation of trees and other objects was not necessary. Thus, we could disregard the vitality parameter. Besides some single trees, the scene

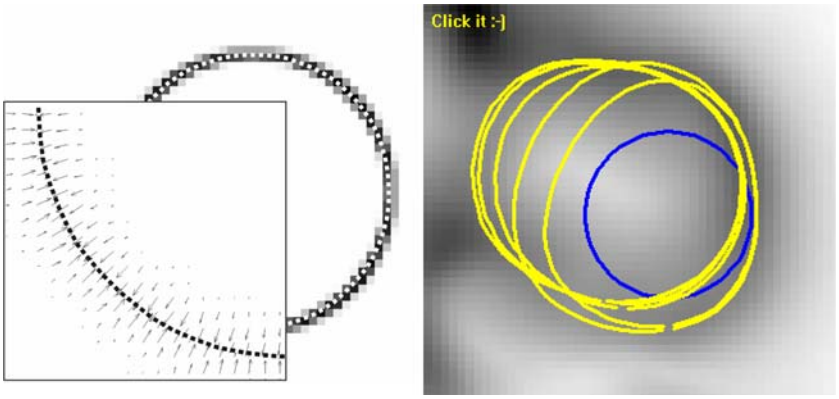

Fig. 10 *Left* External energy for a snake (*background*) and the resulting field of gradients which controls the movement of a snake (*foreground*). *Right* Example for the outline determination with a snake. The *inner circle* shows the initial snake position; the other *closed lines* represent five different optimisation steps

Fig. 11 Examples for different external energy fields: without smoothing (*left*) and with smoothing (*right*). The *grey values* code the amount of external energy: *black parts* represent high values. In the *white areas* the external energy has no influence

contained a few trees touching each other slightly and a number of trees standing in a compact group, as is the usual case in a forest. The *Paris* data set with 62 trees shows parts of the centre of the French capital and was acquired with a digital aerial three-line camera. The DSM was derived by image matching using the same algorithm as for *Grangemouth*. Besides some trees standing in rows, there were a considerable number of compact groups and trees standing rather closely together. Some trees stand close to buildings and thus do not show a crown as circular as in the first two data sets. Furthermore, some of the trees had a few sub-crowns rather than just a single crown. Finally, the *Ravensburg* data set shows a part of this town in the South of Germany. The data were captured using a laser scanner (again using first pulse) with an integrated digital line camera. Since data acquisition occurred in early spring, some of the deciduous trees do not carry leaves, making it impossible to extract them with the described algorithm. The scene contains 108 trees.

The described approach could therefore be tested under a variety of different conditions with respect to scene content, type of depicted trees, data acquisition and ground resolution. The test was carried out with one and the same set of membership functions for size, circularity and convexity. Only the vitality membership function had to be adapted to the different brightness of the image data, since the latter were not radiometrically calibrated.

In order to provide quantitative results, reference data were generated for all four data sets by manually capturing the position and the crown radius of all visible trees. Manual data acquisition was generally performed using the DSM, the image being consulted only in ambiguous cases. The comparison of a manually and an automatically extracted tree was carried out according to the same principle as the overlap determination described in Sect. 3.2.2: both instances were considered as identical if the overlap factor *ov* was greater than 0.5.[5] It should be noted, that these reference data are a type of optimal result of what the approach should deliver from the developers point of view. The relationship between the manually captured reference and the trees in the real world is not discussed here.

An extracted tree is termed True Positive (TP) if an overlapping reference tree can be found, otherwise it is termed False Positive (FP). Trees in the reference data set without correspondence are termed False Negatives (FN). Based on these conventions, the *completeness* C_{om}

[5] We assumed a 1:1 relationship between extracted and reference trees. Thus, if more than one extracted tree was found to correspond to a reference tree, only one extraction result was considered. The inverse case (multiple reference trees corresponding to one extracted tree) did not occur in the test.

and the *correctness* C_{orr} of the extraction result is computed from (Eq. 6)

$$C_{om} = {TP}/{TP + FN} \qquad C_{orr} = {TP}/{TP + FP} \tag{6}$$

In order to characterize the geometric accuracy of the correctly extracted trees, the mean and standard deviation of the distance between the corresponding centres of gravity and the radii were also computed.

4.2 Results

4.2.1 Grangemouth

We first examine two sections of the data set. The results of the first section are depicted in Fig. 12. In the left part we see the DSM and in the middle the automatically extracted results. To the right the reference data are depicted, the latter two superimposed on the intensity channel of the colour-infrared image. Extracted trees are marked by thick white lines. Segments, classified as trees in the selection process but rejected after the refinements using snakes, are depicted as thin white lines. Overall, a completeness of 78% and a correctness of 85% were achieved. In the area marked 01 one can see that while the size of the extracted trees differs somewhat from that of the reference data, the algorithm was able to extract all five of them. In the 02 area two tree hypotheses were initially extracted, but the smaller one, which is partly occluded by the neighbouring larger tree, was rejected in the refinement. The reason is that our object model only consists of single trees and relations between them such as (*partly*) *overlapping* are not considered. The area 03 shows a tree hypothesis, which was correctly rejected during the refinement—the depicted object is actually a hedge, not a tree. Area 04 shows an example of a false positive result while area 05 contains correct, but geometrically imprecise results.

The results of the second window are depicted in a similar way in Fig. 13. The obtained completeness was 81% and the correctness 100%. The trees standing in a row with a decreasing distance between the objects (from top to bottom) are of particular interest. The trees depicted in area 01 are well separated, but the extraction results somewhat overlap, which is a consequence of the difficult snake parameter tuning. The partly occluded tree in area 02 is missed in the extraction process because it is barely visible in the DSM (the manual reference extraction relied on the image in this case). Area 03 again exhibits problems during the refinement stage. The snakes expand too much and thus, a correct tree hypothesis is rejected and the geometric accuracy of the results is rather low.

The results for the complete *Grangemouth* data set are shown in Fig. 14. The extraction results are shown in the left part of the figure and the reference data in the right part. The overall completeness of 72% and the correctness of 78% are somewhat lower than the values reported in the two image sections, but can still be considered useful for many applications.

4.2.2 Hohentauern

As expected the results of the *Hohentauern* data set were not as good as the *Grangemouth* results. We achieved a completeness of only 52%, but a correctness of 94%. The results are depicted in Fig. 15. The left part of the figure shows the DSM, the right part the extraction results. The main problems occurred in the upper left part of the scene. Due to the poor DSM resolution, many trees are barely visible and were not extracted. In the right part of the scene the results are significantly better because the trees are larger (in this part they have a radius of approximately 3 m). Thus, although the algorithm was primarily designed for urban areas, it is in principle also able to extract trees in forests. The main limitation is given by the required high resolution for the DSM—the grid width of 1 m proved to be too coarse for many trees.

4.2.3 Paris

The results for Paris are a little worse than those for Hohentauern. The achieved completeness was 44% and the correctness amounted to 75%. Figure 16 shows (from left to right) the DSM, the intensity channel of the colour-infrared image, and the extraction results. Again, many of the smaller trees could not be extracted due to the low DSM resolution. In addition, trees near buildings did not exhibit a circular shape and a number of trees had more than one crown. These trees were not extracted.

4.2.4 Ravensburg

The *Ravensburg* data set delivered comparable results to the *Paris* one (41% completeness, 66% correctness—see Fig. 17). Besides the low DSM resolution the reason for the rather poor result is the fact that data acquisition occurred in April and, thus, the deciduous trees did not yet carry leaves. Although they are clearly visible in the DSM, they have a relatively small NDVI value and were, thus, not extracted. As can be seen in the right part of the figure, the coniferous trees in the upper right part of the scene were extracted satisfactorily. Disregarding the image data altogether may have delivered better results, but this test was not carried out.

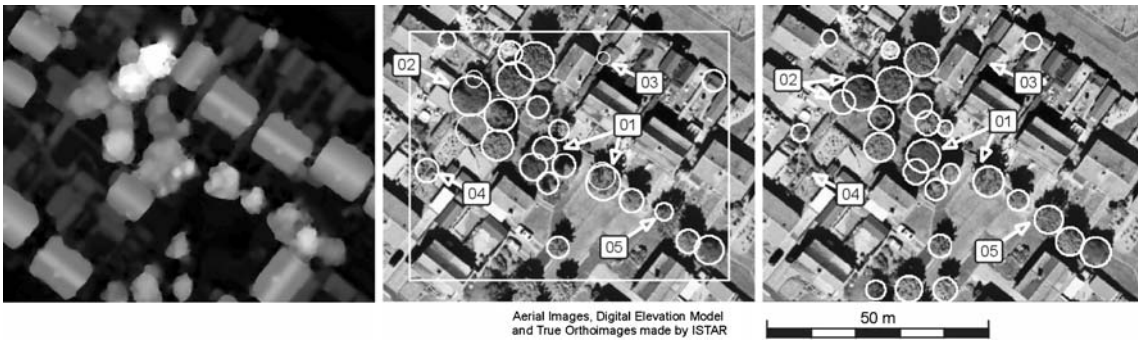

Fig. 12 Project Grangemouth, first section. The DSM is shown in the *left part*, automatically extracted results in the *middle*, and reference data in the *right part*

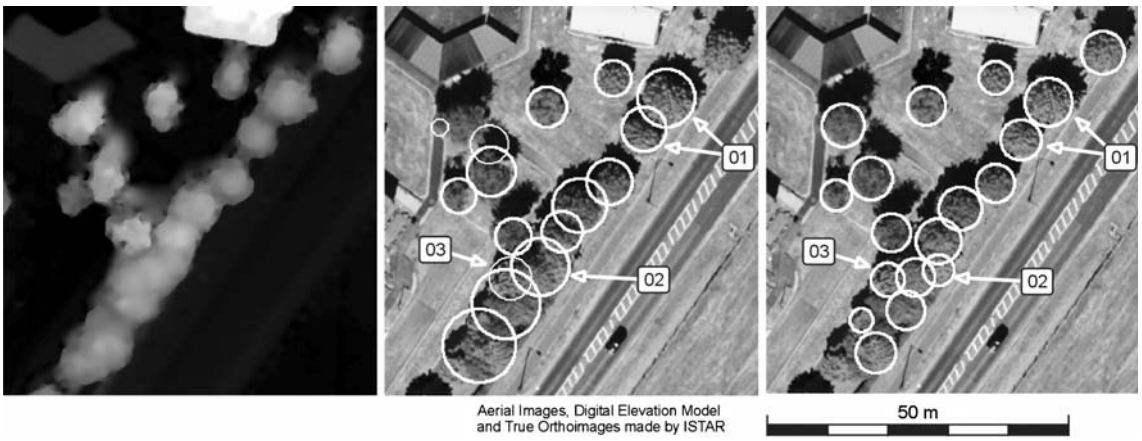

Fig. 13 Project Grangemouth, second section. The DSM is shown in the *left part*, automatically extracted results in the *middle*, and reference data in the *right part*

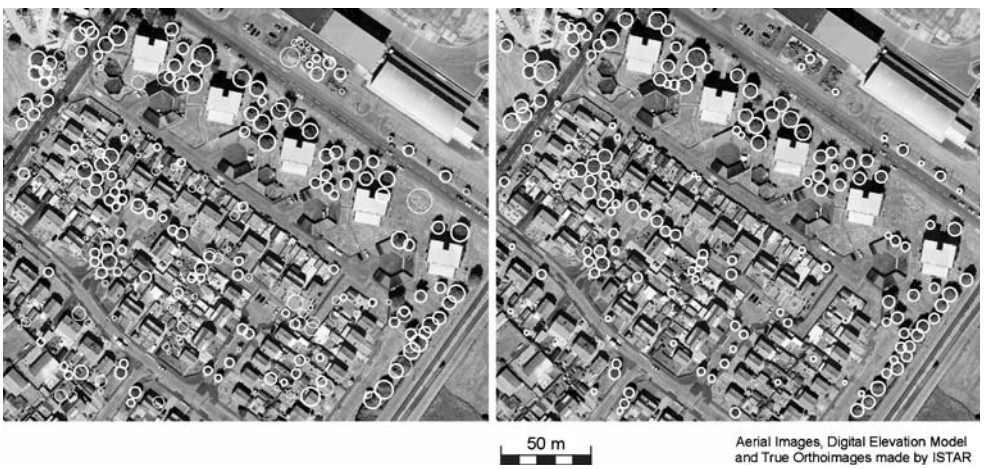

Fig. 14 Project Grangemouth, complete data set: automatically extracted results (*left*), reference data (*right*)

4.3 Discussion of the results

As has been demonstrated the developed algorithm is capable of automatically extracting trees from high resolution remote sensing data in urban areas and in forests. In the latter case, images do not need to be available as the extraction can rely completely on a DSM. For the *Grangemouth* data set with a ground resolution of 0.1 m for the image and 0.2 m for the DSM we achieved a completeness of 72% and a correctness of 78%. For

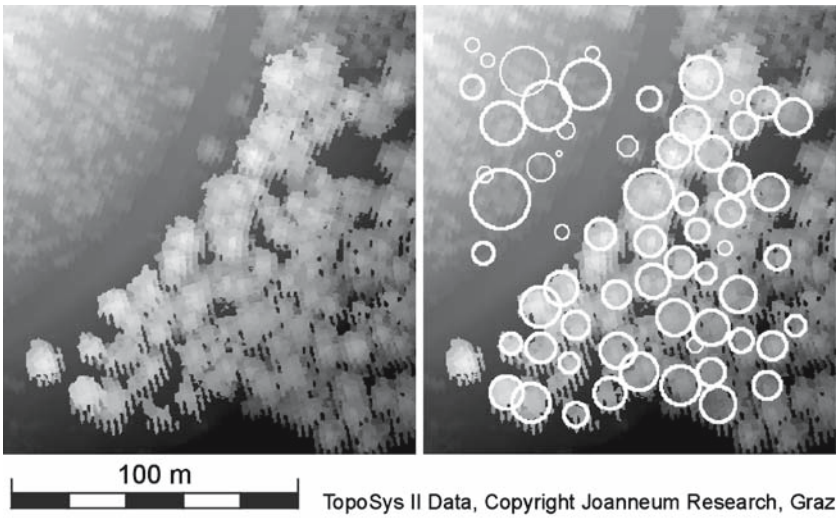

Fig. 15 Hohentauern data set. *Left* DSM. *Right* Extraction results

Fig. 16 Paris data set. *Left* DSM. *Middle* Intensity channel of the colour-infrared image. *Right* Extraction results

Fig. 17 Ravensburg data set. *Left* DSM. *Middle* Intensity channel of the colour-infrared image. *Right* Extraction results

many applications such as visualization, these results are acceptable. The geometric accuracy of the correctly extracted tree position was between 1 and 2 m for all four data sets. For the crown radius the values were between 0.1 and 0.7 m. It is interesting to note that the higher resolution of the *Grangemouth* data did not lead to a better geometric accuracy. This fact can be interpreted such that the definition of a tree position and crown radius, and their extraction from remote sensing data inherently contains an inaccuracy in the range of the reported

values. Thus, the geometric accuracy of the algorithm can be considered to be rather good, which is no surprise given the general state-of-the-art in image analysis: the more difficult problem is to extract the objects, not to accurately delineate them.

The DSM resolution has proved to be the main limiting factor of the approach, a resolution of 1 m significantly deteriorating the results. Other factors influencing the results are the distance of a tree to be extracted from neighbouring objects such as other trees or buildings. A short distance results in violations of our tree model due to partial occlusion or non-circular crowns. Furthermore, trees exhibiting more than one large crown can cause erroneous results, as will deciduous trees carrying no leaves due to the low NDVI.

While the refinement of the results using snakes was successful in a number of cases, we also found some instances where the results became worse. The main reason was a lack of pronounced height differences between the treetop and the borders between the tree and its surroundings. The snakes increase the correctness of the results. At the same time, however, the completeness disimproves somewhat since some of the hypotheses rejected during refinement are in fact correct.

5 Conclusions and Outlook

In this paper, a novel approach for the automatic extraction of trees from co-registered DSMs and colour-infrared imagery is presented. The approach is scale-invariant, i.e. it is free from assumptions about the size of trees to be extracted. Furthermore, we do not make any assumptions about the terrain or the tree heights. This aspect is important because it can be a very difficult task to automatically extract the ground surface and thus the tree height in forest areas. The classification of the hypotheses is based on only four parameters: size, circularity, convexity, and vitality. From these four parameters only the vitality depends on the image material used, the others are object properties.

The approach was applied on four different data sets with the same set of parameters, except for the vitality, in order to demonstrate its general applicability and to investigate its limitations. The results are promising as long as the ground resolution of the DSM is high enough, while coarser input data leads to a deterioration of the quality.

Further developments will focus on the evaluation of the tree hypotheses. The highest potential improvement is expected through a refinement of the membership functions with the help of statistical investigations on large data sets. The mentioned linearization of the NDVI [28] will also be investigated. We will also strive to extend the tree model by including trees with multiple crowns and relationships between trees and neighbouring objects, and to improve the refinement stage of the algorithm. Finally, we wish to investigate possibilities for determining the tree type by extracting more detailed information about the 3D shape of the crown, the fine structure, and the radiometric appearance. Radiometrically calibrated image data will potentially be of significant benefit for this task.

Acknowledgments The European Commission under the contract IST-1999-10510 funded parts of this work. The authors would like to thank the French company ISTAR, which provided the Grangemouth and the Paris data sets. Many thanks go to the Joanneum Research in Graz for the Hohentauern data set, and TopoSys GmbH, Germany, for making available the Ravensburg data set. We are also grateful to the anonymous reviewer for his valuable comments.

References

1. Andersen, H., Reutebusch, S.E., Schreuder, G.F.: Bayesian object recognition for the analysis of complex forest scenes in airborne laser scanner data. In: Kalliany, R., Leberl, F. (eds.) International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXIV, Nr. WG 3A, ISPRS, Graz, pp. 35–41 (2002)
2. Beucher, S.: Watersheds of functions and picture segmentation. In: International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 1928–1931 (1982)
3. Brandtberg, T.: Structure-based classification of tree species in high spatial resolution aerial images using a fuzzy clustering technique. In: The 11th Scandinavian Conference on Image Analysis, Kangerlussueq, 7–11 June 1999, pp. 165–172
4. Brandtberg, T., Walter, F.: Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple scale analysis. Mach. Vis. Appl. **11**, 64–73 (1998)
5. Cohen, L.: On active contour models and balloons. CVGIP: Image Understand **2**(53), 211–218 (1991)
6. Gabet, L., Giraudon, G., Renouard, L.: Construction automatique de modèles numériques de terrain haute résolution en milieu urbain. Bulletin de la Société Française de Photogrammétrie et Télédétection **135**, 9–25 (1994)
7. Gong, P., Sheng, Y., Biging, G.: 3D Model based tree measurement from high-resolution aerial imagery. Photogramm. Eng. Remote Sens. **11**(68), 1203–1212 (2002)
8. Gougeon, F.A.: A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. Can. J. Remote Sens. **3**(21), 274–284 (1995)
9. Gougeon, F., Moore, T.: Individual tree classification using meis-II imagery. In: IGARSS '88 Geoscience and Remote Sensing Symposium, IEEE, vol. 2, pp. 927–927 (1988)
10. Haenel, S., Eckstein, W.: Ein Arbeitsplatz zur automatischen Luftbildanalyse, Mustererkennung 1986, Springer, Heidelberg, pp. 38–42 (1986)
11. Hildebrandt, G.: 100 Jahre forstliche Luftbildaufnahme—Zwei Dokumente aus den Anfängen der forstlichen Luftbildinterpretation. Bildmessung Und Luftbildwesen **55**, 221–224 (1987)
12. Hyyppä, J., Hyyppä, H., Ruppert, G.: Automatic derivation of features to forest stand attributes using laser scanner data.

In: International Archives of Photogrammetry and Remote Sensing, vol. XXXIII, Nr. Part B3, ISPRS, Amsterdam, pp. 421–428 (2000)
13. Hyyppä, J., Hyyppä, H., Letkey, P., Yu, X., Haggren, H., Rönnholm, P., Pyysalo, U., Pitkänen, J., Maltamo, M.: Algorithms and methods for airborne laser scanning for forest measurements. Int. Arch. Photogramm. Remote Sens. vol. XXXVI Part 8/W2, pp. 82–89 (2004)
14. Kass, M., Witkin, A., Terzopoulus, D.: Snakes: active contour models. Int. J. Comput. Vis. **1**, 321–331 (1988)
15. Koenderink, J.: The structure of images. Biol. Cybern. **50**, 363–370 (1984)
16. Larsen, M., Rudemo, M.: Using ray-traced templates to find individual trees in aerial photos. In: Proceedings of the 10th Scandinavian Conference on Image Analysis, Lappeenranta, vol. 2, pp. 1007–1014 (1997)
17. Lillesand, T.M., Kiefer, R.W.: Remote Sensing and Image Interpretation. Wiley, New York, 750 p (1994)
18. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer, Boston, 423 p (1994)
19. Persson, A., Holmgren, J., Söderman, U.: Detecting and measuring individual trees using an airborne laser scanner. Photogramm. Eng. Remote Sens. **9**(68), 925–932 (2002)
20. Pinz, A.: Final results of the vision expert system VES: finding trees in aerial photographs. In: Proceedings ÖAGM 13. Workshop of the Austrian Association for Pattern Recognition, Oldenbourg Schriftenreihe Österreichische Computer Gesellschaft, Wien München, pp. 90–111 (1989)
21. Pollock, R.J.: A model-based approach to automatically locating tree crowns in high spatial resolution images. Image Signal Process. Remote Sens. (eds.) Desachy, SPIE, vol. 2315, pp. 526–537 (1994)
22. Pollock, R.J.: The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model. Dissertation Computer Science, The University of British Columbia, Vancouver, June 1996, 170 p (1996)
23. Schardt, M., Ziegler, M., Wimmer, A., Wack, R., Hyyppä, J.: Assessment of forest parameter by means of laser scanning. In: Kalliany, R., Leberl, F. (eds.) International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXIV, Nr. 3A,ISPRS, Graz, pp. 302–309 (2002)
24. Schneider, S.: Luftbild und Luftbildinterpretation. de Gruyter, Berlin 530 p (1974)
25. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer, Heidelberg, 316 p (1999)
26. Straub, B.: Automatische Extraktion von Bäumen aus Fernerkundungsdaten. Reihe C, Deutsche Geodätische Kommission, München, vol. 572, 99 p (2003)
27. Straub, B., Heipke, C.: Automatic extraction of trees for 3D-City models from images and height data. In: Baltsavias, M., Gruen, A., van Gool, L. (eds.) Automatic Extraction of Man-Made Objects from Aerial and Space Images III. A.A.Balkema Publishers. Lisse Abingdon Exton (PA), pp. 267–277 (2001)
28. Ünsalan, C., Boyer, K.L.: Linearised vegetation indices based on a formal statistical framework. IEEE Trans. Geosci. Remote Sens. **42**(7), 1575–1585 (2004)
29. Weinacker, H., Koch, B., Heyder, U., Weinacker, R.: Development of filtering, segmentation and modelling modules for LIDAR and multi-spectral data as a fundament of an automatic forest inventory system. In: International Archives of Photogrammetry and Remote Sensing, vol. XXXVI Part 8/W2, ISPRS, pp. 50–55 (2004)
30. Winter, S.: Uncertain topological relations between imprecise regions. Int. J. Geograph. Inform. Sci. **5**(14), 411–430 (2000)
31. Wulder, M., White, J., Niemann, K., Nelson, T.: Comparison of airborne and satellite high resolution data for the identification of individual trees with local maxima filtering. Int. J. Remote Sens. **25**(11), 2225–2232 (2004)
32. Zadeh, L.: Fuzzy sets. Inform. Cont. **3**(8), 338–353 (1965)

Author Biographies

Bernd-Michael Wolf (né Straub) received his graduate degree Dipl.-Ing. in geodesy and geoinformatics in 1997, and a Ph.D. degree Dr.-Ing. In 2003, both from Leibniz Universität Hannover in Germany. He is author and co-author of more than 20 international papers, most of them peer-rewiewed. In 2003 he founded SOLVing3D GmbH as a company which develops 3D Machine Vision Solutions based on stereo vision. As President of SOLVing3D he is currently working on new 3D vision solutions for the steel and car manufacturing industry.

Christian Heipke graduated from the Technische Universität München, Germany in 1986 after studies in Geodesy and Geoinformatics at the universities of Hannover, Sydney, and Munich. In 1990 he received a Ph.D. degree (Dr.-Ing.) and in 1994 the venia legendi (Dr.-Ing. habil.), both from Technische Universität München. Throughout 1995 he was a visiting professor at The Ohio State University, Columbus, OH. He subsequently returned to TU München and was awarded another visiting professorship at Ecole Poytechnique Fédérale de Lausanne, Switzerland in spring 1998. In October 1998 he was appointed Head of the Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, where he currently leads a group of about 25 researchers, most of them funded through grants from national and international science organisations and from industry. His professional interests comprise all aspects of digital photogrammetry & remote sensing, image understanding and their connection to GIS.

Christian Heipke has more than 60 refereed publications in scientific journals to his name. He has acted as guest editor and reviewer of major international journals in photogrammetry and remote sensing, and has organized and co-organized a number of international scientific meetings in the fields of photogrammetry, remote sensing and GIS. Currently he serves as chair of the ISPRS working group IV/3 "Automated geo-spatial data acquisition and mapping" and as vice-president of EuroSDR (European Spatial Data Research, formerly known as OEEPE), in which he is responsible for research.