A COMPARATIVE STUDY OF THREE METHODS FOR IDENTIFYING INDIVIDUAL TREE CROWNS IN AERIAL IMAGES COVERING DIFFERENT TYPES OF FORESTS

Mats Eriksson, Guillaume Perrin, Xavier Descombes, Josiane Zerubia

Ariana Research Group CNRS/INRIA/UNSA 2004, route des Lucioles - BP 93 06902 Sophia Antipolis Cedex - FRANCE firstname.lastname@sophia.inria.fr http://www-sop.inria.fr/ariana/

KEY WORDS: Aerial image, forest type, tree crown extraction, segmentation, comparison

ABSTRACT:

Timber volume is an important parameter in forestry and with the position of each individual tree together with its size this parameter can be estimated with higher accuracy than from estimations made at stand level. In this paper, three existing methods for individual tree crown extraction are compared for different types of forests. The types of forests range from planted forests to dense forests. None of the three methods can alone handle all types of forests with satisfactory result. The most obvious remark is that there is a separation between the methods that can handle dense forests and those which can handle sparse forests. None of the three methods compared in this paper can handle both categories in a satisfactory manner.

1. INTRODUCTION

Most of today's silviculture methods are oriented towards optimising the timber volume in the forest. Since many forest companies also care about the nature value, parts of the forest might need to be saved, for instance, to protect a certain habitat. In order to get a good survey of the forest, remotely sensed images are often used. These images are most often manually interpreted in combination with field measurements in order to estimate the forest parameters that are of importance in the decision of how to optimally maintain the forest. Among the most common parameters are the stem number, stem volume, and tree species (Walter, 1998). Interpretations of images are often labour and time consuming. Thus, automatic or semi-automatic methods for interpretation can lower the work load and speed up the interpretation time.

The interpretation is often done using images captured from a far distance from the ground in order to capture as large areas as possible. However, this lower the accuracy of the estimates since the estimation must be done stand wise. Knowledge of the position of each individual tree in the forest together with its size not only increases the accuracy but also makes it possible to better plan the cutting. With this knowledge in mind, research about finding automatic methods for extracting individual tree crowns in aerial or high resolution satellite images has been developed for the last decades.

Different methods for individual tree crown delineation have been presented and different techniques have been used. These techniques include template matching (Pollock, 1996; Larsen and Rudemo, 1998; Olofsson, 2002), region based methods (Pinz, 1989; Culvenor, 2002; Pouliot, King, Bell, and Pitt, 2002; Erikson, 2003; Erikson, 2004; Erikson, 2006), probabilistic methods (Descombes and Pechersky, 2006; Perrin, Descombes, Zerubia, and Boureau, 2006), and contour based methods (Gougeon, 1995; Brandtberg and Walter, 1998).

It is most likely that none of today's methods are alone able to handle all types of forests. Comparative studies of different segmentation methods with different types of forests are therefore of importance in order to clarify how much a method is reliable with respect to a certain type of forest. This knowledge can, for instance, be used to build up an expert system which is supposed to be able to extract individual tree crowns in any type of forests. Comparative studies among the existing methods are rare so far. A comparison between two region growing methods (Erikson, 2003; Erikson, 2004) and a template matching based method (Olofsson, 2002) was made in Erikson and Olofsson, (2005). The comparison was made on plantations of two pine stands and two spruce stands in Sweden. Besides that the results varied among the methods, the conclusion was that none of the methods could capture the correct number of trees in forest (stem diameter > 10 cm) since many trees were not visible in the images due to the fact that they were shaded.

In this paper, an evaluation of three existing delineation methods is made using three images covering different types of forests. The types of forests that are included in the study ranges from equally spaced planted forest, to dense forest, which is naturally regenerated (i.e. the trees have variable size).

2. EXTRACTION ALGORITHMS

2.1 Markov Random Fields

We define a Markov Random Field (MRF) $X = (X_s)$ on the lattice S. A random variable $x_s = (l_s, a_s, b_s, \theta_s) \in \Lambda$ is associated to each site $s \in S$, with $\Lambda = \{0, 1, 2\} \times [a_{min}, a_{max}] \times [b_{min}, b_{max}] \times [0, \pi]$. The labels refer to background pixels $(l_s = 0)$, vegetation pixels $(l_s = 1)$ and tree centres $(l_s = 2)$. An ellipse is associated to each tree which represents its crown. The ellipse is defined by its two axes a_s and b_s together with its orientation θ_s . We embed the problem into a Bayesian framework which consists in maximising the posterior or equivalently the product P(X)P(Y|X), where P(X) is the prior and P(Y|X) is the likelihood associated to the data Y.

To define the prior, we consider pixels s = (i, j) representing tree centres, i.e. those with $l_s = 2$. For such a pixel s, the ellipse representing the tree template is defined as follows:

$$D_1(s) = \left\{ t = (u, v) \in S, f_{(a_s, b_s, \theta_s)}(t) = \frac{U^2}{a_s^2} + \frac{V^2}{b_s^2} \le 1 \right\},$$
(1)

where:

$$U = (u - i)\cos\theta_s + (v - i)\sin\theta_s$$
$$U = (u - i)\sin\theta_s - (v - i)\cos\theta_s$$

We define the tree neighbourhood as follows:

$$D_2(s) = \left\{ t = (u, v) \in S/D_1(s), f_{(a_s + e, b_s + e, \theta_s)}(t) \le 1 \right\},$$
(2)

where e is a parameter defining the neighbourhood size. We then define some interactions which favour label 1 inside the templates and label 0 in their neighbourhood. Besides, templates overlapping is penalised. The resulting potential is written as follows:

$$\forall t \in D_1(s), V_s(t) = \beta_a \frac{1 - f_{(a_s, b_s, \theta_s)}(t)}{\sum_{r \in D_1(s)} 1 - f_{(a_s, b_s, \theta_s)}(r)} sgn_1(l_t)$$

+ $\beta_r \delta_{l_t=2}$ (3)

$$\forall t \in D_2(s), V_s(t) = \beta_a \frac{f_{(a_s, b_s, \theta_s)}(t) - 1}{\sum_{r \in D_2(s)} f_{(a_s, b_s, \theta_s)}(r) - 1} sgn_2(l_t)$$

where $sgn_1(0) = 1$, $sgn_1(1) = -1$, $sgn_1(2) = 0$ and $sgn_2(0) = -1$, $sgn_2(1) = 1$, $sgn_2(2) = 0$. To define the likelihood, we assume the data Y is independently distributed conditionally on X and Gaussian (the data distribution for tree centres and vegetation classes is identical). The Gaussian distribution parameters, (μ_0, Σ_0) and $(\mu_{1,2}, \Sigma_{1,2})$, are empirically estimated from the result of a K-means algorithm. We then have:

$$\forall s \in S, p(y_s | x_s = l) = \frac{1}{(2\pi det(\Sigma_l))^{3/2}} \times \exp\left\{-\frac{||(y_s - \mu_l)^t \times (y_s - \mu_l)||_{\Sigma_l}}{2}\right\}.$$
 (4)

The posterior is then written as follows:

$$P(X|Y) \propto \exp \left[\sum_{s \in S: l_s = 2} \sum_{t \in D_1(s) \cup D_s(s)} V_s(t) + \sum_{s \in S} p(y_s|l_s)\right].$$
 (5)

The optimisation is performed using a Metropolis dynamics embedded in a simulated annealing scheme. Further details can be found in Descombes and Pechersky, (2006)

2.2 Marked Point Processes

The Marked Point Processes (MPP) method consists in modelling the stands as configurations of geometric objects (the trees are represented by ellipses or ellipsoids). These objects are defined on a state space $S = \mathcal{P} \times \mathcal{K}$ by their location and their mark (i.e. geometric attributes). The associated marked point process X is a random variable whose realisations are random configurations of objects. This approach yields an energy minimisation problem, tackled with a Reversible Jump Markov Chain Monte Carlo algorithm and simulated annealing (Perrin, Descombes, and Zerubia, 2005). The energy of the process $U(\mathbf{x})$ embeds a regularisation term (prior density $U_p(\mathbf{x})$), which introduces some interactions between the objects, and a data term $U_d(\mathbf{x})$, which links the objects to the features to be extracted :

$$U(\mathbf{x}) = U_p(\mathbf{x}) + U_d(\mathbf{x}) \tag{6}$$

The prior energy $U_p(\mathbf{x})$ gathers the knowledge we have about the configuration we would like to extract. This includes prior information about the location or the distribution of the marks of the objects for example, but also interactions between a group of objects. We cope with Gibbs processes with pairwise interactions (cliques of order 2), by taking into account a repulsion between overlapping objects (Strauss process), and an attraction to special patterns such as alignments in plantations (see Perrin, Descombes, Zerubia, and Boureau, (2006) for more details).

The data energy $U_d(\mathbf{x})$ is calculated at the object level :

$$U_d(\mathbf{x}) = \gamma_d \sum_{x_i \in \mathbf{x}} U_d(x_i) \tag{7}$$

where $\gamma_d > 0$ and $U_d(x_i) \in [-1, 1]$. An object will be attractive and therefore favoured if its data energy term is negative. On the contrary, a positive data energy will yield a repulsion for the related object. In practise, we favour a difference of infrared value distribution between the pixels inside the object and the pixels in the shadow area (all around the tree in the 2D model with ellipses in dense areas, or just in the direction of the sun in the 3D model with ellipsoids in sparse vegetation), because high reflectance in the near infrared and the shadow are the two main criteria to extract the trees (see Perrin, Descombes, and Zerubia, (2006) for further details).

2.3 Region growing

The region growing method (RG) needs seed points for each region before growing. To find these points, the first band of the image is thresholded with a chosen threshold (20% of max) and a distance transform (Borgefors, 1986) is performed on the resulting image. The resulting distance image is smoothed with a Gaussian filter and local maxima in the smoothed image are found. These maxima constitute the seeds.

Each seed is then grown to become a region. The order each pixel connects to a region is decided by the pixel value if the pixel is a border pixel of a region. If it is not a border pixel it has to wait until that is the case before it is connected to a region. The higher the value of the border pixel the faster it is connected to a region.

However, the pixel values are not taken from the original image, f. Instead f is transformed into a new image, called I_{rw} , which is the image used in the growing step. The transformation is made by simulating "random walks" for each seed point. Each pixel value of I_{rw} represents the number of times the simulated particles have reached the pixel. The particles are only allowed to move inside a certain neighbourhood, N(p), of the current position, p. At time t = 0, the position of one particle is a seed point and at time t = 1, the particle is at a position $q \in N(p)$ and thus the pixel value for $I_{rw}(q)$ is increased by one. The size of the neighbourhood, N(p), depends on the sum of the included grey levels (from f) in the neighbourhood. Thus, if the sum is less than a certain threshold the neighbourhood is grown by one pixel in each direction until the sum exceeds the threshold.

The position $q \in N(p)$ is selected randomly and each pixel in N(p) has the same probability to become the next position. However, the particle may not be allowed to move to q due to the fact that another random number, z, uniformly chosen in [0, N], where N is the number of different grey levels, is larger than the grey level at q, i.e. if z < f(q) the move is taken otherwise randomise again. The last constrain prevents the particle to move to a dark background. A fixed number of iterations, N_{iter} , for each seed point is used to construct I_{rw} . Further details can be found in Erikson, (2006). This method produces different resulting images compared to the other two methods. Instead of having a position and a size of each tree crown (an ellipse) as for the MRF and MPP methods, RG techniques produce regions with contours corresponding to the visible tree crowns in the image. In order to make the comparison between this method and the other two, the segments are converted into points representing the crown centre by calculating the centre of mass of the segment. As the radius of the tree crown the maximum distance from the centre of mass to all other points in the segment is used.

3. RESULTS AND DISCUSSION

Three images were used in this study. The first image, *plantation*, contained equally spaced planted forest, the second image, *dense*, contained a dense forest with one species while the last image, *mixed*, covered a dense naturally regenerated forest with mainly four species and with tree crowns of different sizes. The first two images captured French forests (with 50 cm resolution) while *mixed* captured Swedish forests (with 10 cm resolution). All images were captured from an aircraft with an analog camera using a colour infrared film. Even though all the three methods used here can produce an estimation of the sizes of the tree crowns only the locations of the trees were used in the comparison.

Results on *plantation* are shown in Fig. 1 for MRF, in Fig. 2 for MPP, and in Fig. 3 for RG. The positions of the tree crowns are shown by white crosses which overlaid the original image.



Figure 1. 934 positions (white crosses) found by MRF for *plantation*. © IFN/INRIA.



Figure 2. 865 positions (white crosses) found by MPP for *plantation*. © IFN/INRIA.



Figure 3. 1023 positions (white crosses) found by RG for *plantation*. (c) IFN/INRIA.

For the *plantation* the three methods found more or less all the tree crowns. What differed was that RG had more difficulties with non-forested areas than the other two methods. The shape template embedded into the MPP and MRF avoids false alarms in vegetation areas, where the radiometry is close to the radiometry of the trees, by the lack of shadows in the vegetation areas.

Fig. 4 shows the result from *dense* using MRF, Fig. 5 shows the result using MPP, while Fig. 6 shows the result using RG. Again, the positions of the tree crowns are shown by white crosses which overlaid the original image.



Figure 4. 376 positions (white crosses) found by MRF for *dense*. © IFN/INRIA.

In *dense* the results differed quite a lot. Both MRF and MPP had tendencies to underestimate the number of tree crowns in the real dense areas. This is due to both the lack of shadows and the shape template, which does not match the tree geometry in this case. On the other hand, RG seemed in some cases to overestimate the number of crowns due to bad seeds.

The results from *mixed* are shown in Fig. 7 for MRF, in Fig. 8 for MPP, and in Fig. 9 for RG. Recall that the resolution is different for this image than for the other ones (10 cm instead of 50 cm). The results shows that MRF found too many tree crowns, many of them corresponding to subcrowns or branches of tree crowns.



Figure 5. 673 positions (white crosses) found by MPP for *dense*. © IFN/INRIA.



Figure 7. 494 positions (white crosses) found by MRF for *mixed*. © CBA/INRIA



Figure 6. 800 positions (white crosses) found by RG for *dense*. © IFN/INRIA.

RG and MPP gave similar results as long as the crowns were not connected to each other. In the most dense cases the result was similar as for *dense*, that is, MPP estimated the crown location as being in the outer part of the real crown. Further more, MPP had a tendency to miss the smallest tree crowns which RG found.

4. CONCLUSIONS AND FUTURE WORK

None of the three delineation methods could alone handle all the types of forests used for this comparison. With the results in mind, the guidelines are that if a forest is so sparse that ground patches are present in between the crowns one should use MPP as the best extraction method among the three. For planted forests it does not matter which method is chosen as long as only the loca-



Figure 8. 328 positions (white crosses) found by MPP for *mixed*. © CBA/INRIA

tion of the crowns matters, i.e. neither the computer time is taken into account here nor the non-forested areas around the forest that caused the problem for RG in this case. For dense forests the best choice is to use RG.

However, there can be problems if an image captures a forest which has different stands according to the sparseness/denseness. In the most difficult cases, such as when there are both sparse stands with lots of space in between the tree crowns and dense stands in the same image, it is necessary to do a pre-segmentation of the forest. The pre-segmentation should then divide the forest into different stands according to the sparseness/denseness before it is possible to apply the best delineation method for each type of stand. Thus, further comparison between the existing delineation methods must be done as well as developing segmentation methods for stand delineation together with classification of each stand according to sparseness/denseness.



Figure 9. 424 positions (white crosses) found by RG for *mixed*. (c) CBA/INRIA

ACKNOWLEDGEMENTS

This work was partially founded by ARC MODE de VIE collaborative project at INRIA. The authors would like to thank the French Forest Inventory (IFN) and the Centre for Image Analysis (CBA) in Uppsala (Sweden) for providing the data.

REFERENCES

Borgefors, G., 1986. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing* 34, 344–371.

Brandtberg, T. and Walter, F., 1998. Automatic delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis. *Machine Vision and Applications* 11(2), 64–73.

Culvenor, D. S., 2002. Tida: an algorithm for delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geoscience* 28(1), 34–44.

Descombes, X. and Pechersky, E., 2006. Tree crown extraction using a three states markov random field. *manuscript in progress*.

Erikson, M., 2003. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research* 33(8), 1557–1563.

Erikson, M., 2004. Species classification of individually segmented tree crowns in high-resolution aerial images using radiometric and morphologic image measures. *Remote Sensing of Environment* 91(3), 469–477.

Erikson, M., 2006. Two preprocessing techniques based on grey level and geometric thickness to improve segmentation results. *Pattern Recognition Letters* 27(3), 160–166.

Erikson, M. and Olofsson, K., 2005. Comparison of three individual tree crown detection methods. *Machine Vision and Applications* 16(4), 258–265.

Gougeon, F. A., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing* 21(3), 274–284.

Larsen, M. and Rudemo, M., 1998. Optimizing templates for finding trees in aerial photographs. *Pattern Recognition Letters* 19, 1153–1162.

Olofsson, K., 2002. Detection of single trees in aerial images using template matching. In *ForestSat 2002, Operational tools in forestry using remote sensing techniques.*, Edinburgh, Scotland.

Perrin, G., Descombes, X., and Zerubia, J., 2005. Point processes in forestry : an application to tree crown detection. Research Report 5544, INRIA.

Perrin, G., Descombes, X., and Zerubia, J., 2006. A Non-Bayesian Model for Tree Crown Extraction using Marked Point Processes. Research Report 5846, INRIA. Perrin, G., Descombes, X., Zerubia, J., and Boureau, J. G., 2006. A Non-Bayesian Model for Tree Crown Extraction using Marked Point Processes. In *International Precision Forestry Symposium*, University of Forestry, Stellenbosch, South Africa.

Pinz, A., 1989. Final results of the vision expert system VES: Finding trees in aerial photographs, pp. 90–111. OCG-Schriftenreiche 49, Oldenbourg Verlag.

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

Pouliot, D. A., King, D. J., Bell, F. W., and Pitt, D. G., 2002. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment* 82, 322–334.

Walter, F., 1998. Fjärranalys för skoglig plannering. (summary: Remote sensing for forestry planning.). Report 9, The forestry Research Institute of Sweden, Science Park, SE-751 83 Uppsala, Sweden. (In Swedish with English summary).