Can. J. Remote Sensing, Vol. 31, No. 3, pp. 255–267, 2005

Approaches for optimal automated individual tree crown detection in regenerating coniferous forests

Darren Pouliot and Doug King

Abstract. Automated tree detection provides a means to acquire information on tree abundance and spatial distribution, both of which are critical for evaluating the status of regenerating forests. It is also often a precursor to automated tree delineation, which typically utilizes image data surrounding a detected crown point. However, obtaining consistently accurate detection results has proven difficult owing to errors associated with image scale. In this paper, four approaches that reduce this scale dependence are evaluated, including (1) determination of optimum global image smoothing to apply predetection, (2) determination of optimum local image smoothing to apply predetection, (3) determination of the optimal local window size for use in the detection algorithm, and (4) post-detection merging of initially defined crown segments. Each approach was applied to three datasets acquired by different sensors and with different regenerating forest conditions. A common local maximum tree detection algorithm was implemented for approaches 1–3, and a watershed segmentation algorithm was applied in approach 4. Detection accuracy was evaluated using standardized methods. The highest accuracies for each dataset were obtained with approaches based on local scale representations where the regenerating structure favored such approaches. However, more consistent accuracies across all datasets were obtained with the optimum global scale approach. Post-detection merging of adjacent crown segments produced the poorest results. Error sources and the advantages and disadvantages of each approach are discussed in terms of developing more operational methods for automated tree detection in regenerating forests.

Résumé. La détection automatique des arbres constitue un moyen d'acquérir une information sur l'abondance et la répartition spatiale des arbres, deux facteurs critiques pour l'évaluation du statut des forêts en régénération. Cette dernière est souvent un pré-requis à la délimitation automatique des arbres individuels, qui utilise habituellement les données image autour d'une pointe de couronne détectée. Toutefois, l'obtention de résultats de détection toujours précis s'est avérée difficile dû aux erreurs associées à l'échelle de l'image. Dans cet article, quatre approches réduisant cette dépendance à l'échelle sont évaluées, incluant (1) la détermination du lissage global optimal pour application à la pré-détection, (2) le lissage optimal local de l'image pour application à la pré-détection, (3) la détermination de la dimension optimale de la fenêtre locale pour utilisation dans l'algorithme de détection et (4) la fusion post-détection des segments de couronnes définies initialement. Chacune des approches a été appliquée à trois ensembles de données acquis par différents capteurs et dans des conditions différentes de régénération forestière. Un algorithme commun de détection d'arbres basé sur le maximum local a été implanté pour les approches 1–3, alors qu'un algorithme de segmentation de bassin versant a été appliqué à l'approche 4. La précision de détection a été évaluée à l'aide de méthodes standard. Les taux de précision les plus élevés pour chaque ensemble de données ont été obtenus à l'aide des approches basées sur les représentations à l'échelle locale, où la structure de régénération favorisait de telles approches. Toutefois, des précisions plus constantes pour l'ensemble des données ont été obtenues avec l'approche basée sur l'échelle globale optimale. La fusion post-détection des segments de couronnes adjacentes a produit les résultats les plus faibles. On discute des sources d'erreur et des avantages et désavantages de chacune des approches en terme du développement de méthodes plus opérationnelles pour la détection automatique des arbres dans les forêts en régénération.
[Traduit par la Rédaction]

Introduction

In regenerating forests, accurate knowledge of tree abundance and spatial distribution is important in predicting future forest conditions and determining if these conditions are sufficient to meet the preset management objectives for the site (Pitt et al., 1997). High spatial resolution remote sensing provides a potentially low cost means to obtain this information if efficient and reliable methods for detection and measurement of individual trees in imagery can be developed. Automated individual tree detection algorithms have been developed for this task, but accuracy has proven to be highly dependent on scale, making consistent results difficult to achieve (Gougeon, 1995; Dralle and Rudemo, 1996; Brandtberg and Walter, 1998; Gougeon and Leckie, 1999; Pinz, 1999; Walsworth and King, 1999; Wulder et al., 2000; Culvenor, 2000; Pitkänen, 2001; Pouliot et al., 2002; Erikson, 2003). Most algorithms, with the

Received 23 June 2004. Accepted 22 November 2004.

D. Pouliot and D. King.[1] Department of Geography, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada.

[1]Corresponding author (e-mail: doug_king@carleton.ca.).

exception of those of Gougeon and Leckie (1999) and Pouliot et al. (2002), have been tested in mature forest conditions. However, common to these studies has been a noted difficulty of defining a scale representation in a consistent manner to achieve optimal detection results. Here, the term "scale" refers to the "grain" or detail of an image and not the extent or coverage.

Considering an image with a globally defined scale, small trees relative to the scale are often not detected (omission error) because of their weak spectral and spatial response in the imagery, whereas larger trees may cause false detection (commission error) owing to crown branches being detected as single tree crowns. In image acquisition, a common approach is to acquire imagery with a pixel size suitable to detect the smallest required object (i.e., using a criterion such as half the object size), as this minimizes omission error. This can lead to high commission errors in larger crowns, however, because of within-crown branch clusters being detected as single crowns and image noise. An "optimal" global scale is defined as that which minimizes both errors of omission and errors of commission, producing the best overall detection results. An alternative approach is to find the optimum local scale for each object of interest (i.e., trees). Optimum local scale is not defined for a minimized error but is the scale where no error exists. That is, the appropriate local scale can be defined so that the tree is detected as a single object.

The scale of an image can be modified in several ways. Image smoothing is most commonly applied, with the appropriate smoothing factor or window sample size criteria being difficult to define for optimal tree detection. Thresholds or parameters can also be considered as scale modifications when linked directly to object size or indirectly to size through an image property that is size related. For example, the variance threshold in a region-growing algorithm indirectly controls the size of the region found.

The scale component of automated tree detection analysis has only been addressed in a few studies concerned with mature forests. Brandtberg and Walter (1998) made use of scale space theory to constrain local maximum detection based on multiscale edge detection in scanned colour infrared photography and more recently for local based scale optimization in lidar imagery (Brandtberg et al., 2003). The scale space approach involves examination of changes in image brightness with changes in scale simulated by a Gaussian smoothing function (Lindeberg, 1994). Its advantage is that it does not require any prior information regarding the objects to be extracted. However, the number of objects to be extracted from the set of scale space objects and the scale interval and range must be selected and are typically determined arbitrarily or empirically using field data. Further, high-contrast objects are likely to be found as significant objects, regardless of whether they represent actual image features of interest (Hay et al., 2003). In cases where prior knowledge exists about the objects of interest, improved object extraction and simplification of the scale selection process are possible. For example, Wulder et al. (2000) exploited the spatially dependent nature of tree crowns using semivariance analysis to estimate the optimal window size for local maximum filtering but found that using local edge detection produced better window size estimates. Pouliot et al. (2002) also found that the latter approach improved tree detection results in young regenerating forests. These optimal window size approaches can be effective but are not ideal, as only one local maximum is returned within the window. For larger windows, more than one maximum may exist, representing several trees, resulting in omission errors. In a similar approach, Pinz (1999) used the mean brightness value in successively larger sample windows to evaluate whether a given local maximum was likely to represent a tree crown. This approach removes error associated with large windows containing multiple crowns, but how the evaluation was performed and its effects on the final results were unclear. Culvenor (2000) proposed a simple method to determine the optimal global scale. It is based on an examination of the number of local maxima detected at different levels of scale simulated using a Gaussian filter, followed by selection of the optimum global scale as the point at which the rate of change of the second derivative is a maximum. In lidar data, Persson et al. (2002) used local surface fitting to determine which of three user-defined scales was optimal for detection of a given tree crown.

Each of these approaches has been found to improve results in their intended applications. However, comparison and adaptation of these methods to other studies is difficult for several reasons. First, a wide range of accuracy evaluation methods are applied, and therefore results are not directly comparable. Second, methods have been developed and tested with specific image formats and forest conditions, making it difficult to evaluate their dataset dependence. Third, methods vary in terms of required parameters and the sensitivity of tree detection to these parameters. This makes it difficult to determine how precisely such parameters must be specified and how easily this precision can be achieved. Lastly, the methods are often presented as part of a larger study involving automated detection and delineation, and in some cases the methods used to optimize tree detection are not well described.

The purpose of this study was to determine the best approach for automated tree crown detection in regenerating forests. The issues identified previously are explicitly addressed in the selection of data types and analytical procedures. Four approaches are compared that integrate methods previously presented in the literature with new methods. They are evaluated in terms of their accuracy using standard methods, their consistency across a range of data types and regeneration conditions, and their input parameter dependence. The four approaches are as follows: (1) the local maxima scale relation described by Culvenor (2000), (2) the variable window size method described in Wulder et al. (2000) and modified by Pouliot et al. (2002), (3) a new approach for local image smoothing based on variable window size methods, and (4) a new approach based on post-detection merging of detected segments. The first three are applied predetection to identify an optimal representation of scale for local maximum based

detection analysis. The last approach involves evaluation of detection and initial crown segmentation results to identify segments that should be merged to represent a single crown.

Datasets

Three subsets of larger datasets were used consisting of a range of variability in regeneration tree structure that would be typical of operational applications. The datasets consisted of an experimental study with highly controlled conditions (ESS), a young regenerating cutover (YCO), and an older regenerating cutover (OCO). The latter two were from operational forest management areas. As each dataset was from separate and larger studies of the Ontario Ministry of Natural Resources (OMNR), the Canadian Forest Service (CFS), and Carleton University, image types and pixel sizes varied. Specific characteristics of each dataset are listed in **Table 1**. Example images from each dataset are shown in **Figure 1**. All imagery was acquired before deciduous and herbaceous vegetation leaf flush to minimize commission error from background vegetation.

Experimental study site (ESS)

The ESS dataset was used in Pouliot et al. (2002) in initial development of a detection–delineation algorithm. It was compiled from an experimental study site established in 1994 by the OMNR and CFS as part of an effort to identify the effects of various levels of vegetative competition on black spruce (*Picea mariana*) and jack pine (*Pinus banksiana*) crop trees (Bell et al., 2000). Trees were planted at a spacing of 1 m in 7 m × 7 m plots with different species and intensities of competing vegetation. For this study, only plots containing black spruce trees were used because by 2000 the jack pine crowns had grown into one another in such a way that individual detection was not possible. Plots with competing vegetation were also excluded to remove their effects on automated detection.

Images of the study site were acquired in a leaf-off condition on 25 April 2000 between 1200 and 1430 hours using a Kodak DCS 460 CIR digital camera with ~5 cm nominal ground pixel size. Three spectral bands were acquired in the green (500–600 nm), red (570–780 nm), and near infrared (710–800 nm). The camera uses a single charged coupled device (CCD) with a filter corresponding to the selected wavelengths placed over the photograph sites in a Bayer pattern. As each photograph site receives radiance from one spectral band, interpolation is used to derive data for the other two bands.

For validation of detection results, tree positional data were not required, as the initial planting arrangement of 1 m spacing was known and was clearly identifiable in the imagery.

Young cutover (YCO)

The YCO dataset represented young regeneration conditions occurring after harvesting and planting. Black spruce and jack pine were planted in 1997, but most of the pine was natural infill. Competition consisted of pockets of high-density aspen (*Populus tremuloides*) and lower density white birch (*Betula papyrifera*). The aspen tended to be considerably taller than the birch, and both aspen and birch often overtopped the spruce and pine trees. The topography was hilly with gentle slopes.

Image data were acquired on 11 May 2002 between 1130 and 1500 hours using a DuncanTech MS3100 CIR digital camera, which employs beam-splitting technology to focus incoming light on three CCDs and capture separate wavelengths simultaneously. This allows it to produce higher spectral quality image data (i.e., non-overlapping bandwidths and fewer correlated bands) than can be acquired using single CCD cameras such as the Kodak DCS. Images were acquired in the green (500–600 nm), red (600–700 nm), and near infrared (700–900 nm), with ~6 cm nominal ground pixel size.

For validation, circular plots with a radius of 3 m were laid out in the field, and the plot centre was marked using a 40 cm × 40 cm white board set out before imagery was acquired. The direction and distance of each conifer tree within the plot to the plot centre were measured. These data were used in a geographical information system (GIS) to create a reference map of crown locations.

Old cutover (OCO)

The OCO study site was part of the OMNR Fallingsnow Ecosystem Project designed to evaluate the effects of chemical and mechanical vegetation competition on flora and fauna (Lautenschlager et al., 1997). It represented older regeneration conditions occurring after harvesting and planting. White spruce (*Picea glauca*) was planted in 1988, but balsam fir (*Abies balsamea*), white pine (*Pinus strobus*), jack pine, and white cedar (*Thuja occidentalis*) had naturally infilled. Competition consisted mostly of aspen and alder (*Alnus* spp.). Aspen ranged from low to high density and was typically taller than conifers. Alder also ranged from low to high density but tended to be of equal or lower height than conifers. The site had

Table 1. Summary of selected characteristics for datasets used in the study.

Study site[a]	Pixel size (cm)	Max. view angle (°)	No. of trees	Dominant species	Age (years)	Crown diameter (cm)			Average height (cm)	Density (trees/ha)
						Mean	SD	Range		
ESS	5	7	197	Spruce	6	85	27	16–145	127	10 000
YCO	6	10	265	Spruce, pine	5	58	31	5–135	110	3 155
OCO	15	25	144	Spruce	15	129	43	40–230	>200	2 300

[a]ESS, experimental study site; OCO, old cutover; YCO, young cutover.

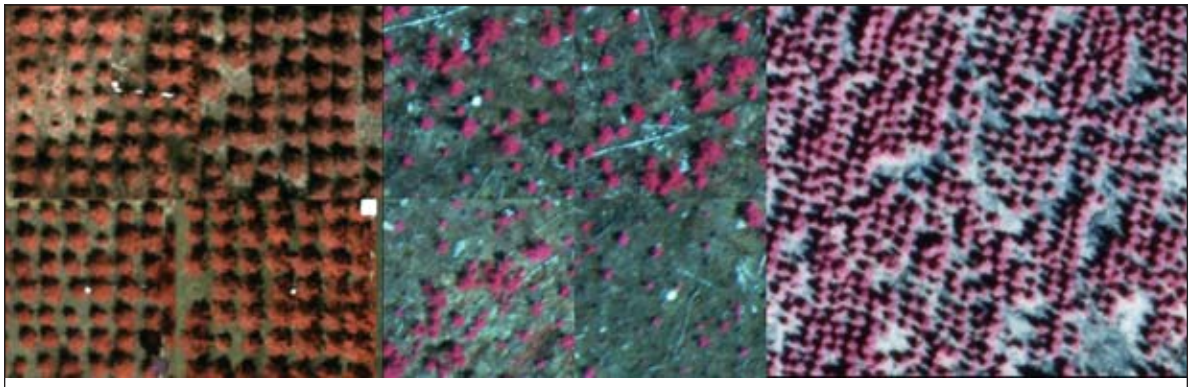

Figure 1. Example images of datasets used in the study: (left) experimental study site (ESS); (middle) young cutover (YCO); (right) older cutover (OCO).

moderate to steeply sloped topography, but data used in this study were only from the moderately sloped areas.

Colour infrared photography was acquired at 1 : 5000 scale on 21 May 2003 between 1100 and 1500 hours using a Wild RC-10 aerial mapping camera with Kodak Aerochrome Infrared II 2443 film. The photographs were scanned at 850 dpi to 15 cm nominal ground pixel size. These data represent typical operational imagery used in many forestry applications and were acquired as part of the Fallingsnow project to provide detailed maps of conifer and other vegetation cover for wildlife habitat and population studies.

For validation, seven 1 : 500 scale colour photographs were acquired and scanned at 400 dpi to give a nominal ground pixel size of 3 cm. Visual interpretation of these photographs was carried out by the first author for validation of detection accuracy with this dataset.

Methods

Image preprocessing

Before automated tree detection is performed, most studies have applied some means of initial separation of vegetation pixels from nonvegetation pixels. This serves to eliminate brightness gradients not associated with vegetation, which could result in false detection of brightness peaks as trees. Thresholding is commonly employed for this task (Gougeon, 1995; Gougeon and Leckie, 1999; Walsworth and King, 1999; Culvenor, 2000; Pitkänen, 2001; Pouliot et al., 2002; Persson et al., 2002; Erikson, 2003), but in this study supervised maximum likelihood classification of five classes (soil, water, snow, shadow, and vegetation) was used for the following reasons. First, vegetation pixels in the imagery were spectrally distinct from all other classes (minimum Bhattacharrya separability of 1.999), suggesting the potential for high classification accuracy. Second, because these classes were distinct, the classification results were less dependent on well-defined training data, thus simplifying the classification task and increasing its repeatability. Third, statistical classifiers are less sensitive to image brightness variations such as that resulting from bidirectional reflectance or topography, as the defined statistical criteria provide greater discretion in pixel class assignment compared to thresholding.

The near-infrared band was used for automated detection in the ESS and YCO datasets and the green band was used for the OCO dataset, as vegetation was saturated in some portions of its near-infrared imagery. An initial Gaussian filter (1σ, where σ is the standard deviation in pixels) was applied to reduce image noise and convert the data to floating-point format so that local maxima were represented by a single pixel. In byte format, local maxima often exist as groups of pixels due to the reduced precision. In very high resolution data, such as those used in this study, considerable presmoothing can be applied for noise removal without drastically impacting the image morphology. With coarser resolution imagery, however, care must be taken because smoothing data may negatively affect detection accuracy.

Local maxima detection was carried out using a moving 3 × 3 window where a local maximum was identified only when the centre pixel in the window was greater than all eight of the surrounding neighbours. This is the most common method applied in tree detection studies and was used for the three optimum image scale approaches described in the next section. The fourth approach used a watershed segmentation method for local maxima detection. The detection results were the same using either method, but in the watershed method the local minima network surrounding a local maximum was also defined, providing an initial crude segmentation of the crowns.

Approaches to optimal detection

Number of local maxima and smoothing relation (LMSR)

The number of local maxima and smoothing relation (LMSR) approach to optimal detection attempts to identify the optimal level of global smoothing to apply before local maxima detection is carried out. It is adopted from Culvenor (2000), who applied it to simulated imagery of mature eucalyptus forests. In this approach, the relation between the number of local maxima detected and the Gaussian smoothing factor is evaluated to identify the optimum. The Gaussian filter is given as follows:

$$f(x) = \exp[-(x)^2/2\sigma^2]/2\pi\sigma^2 \quad (1)$$

where x is the distance from the filter centre, and σ is the smoothing factor represented as the standard deviation in pixels.

The local maxima smoothing relation is determined for the original image scale and for simulated scales derived using a range of σ values. For the three datasets used in this study, σ ranged from 0.1 to 4.0 in increments of 0.2. The automated method given in Culvenor (2000), which determined the optimal scale as the maximum rate of change in the second derivative, proved to be highly sensitive to noise, giving inconsistent results. For this reason, the relation was visually assessed to determine the optimal smoothing factor. Referring to **Figure 2**, an example smoothing relation, the first step in the visual assessment is to plot the relation as the number of local maxima versus the smoothing factor. The next step is to determine the longest line that can be fit to the curve (plotted relation), starting with the last point on the right side of the graph and moving towards the first point on the left side while maintaining a suitable fit. This is defined as a linear fit, where residuals are relatively small and equally distributed on either side of the line. The final step is to determine the point where this line and the curve start to diverge. This point is taken as an indication of the optimal smoothing factor required. **Figure 2** shows the line that best fits the example smoothing relation and identifies the optimal Gaussian smoothing factor.

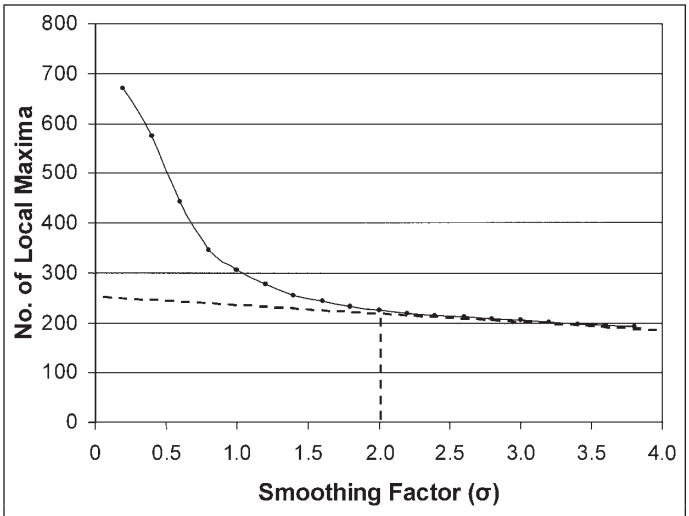

Figure 2. An example smoothing relation. The solid line indicates an example local maxima smoothing relation, and the broken line the visual rule used to determine the optimal Gaussian smoothing factor.

Variable local window size (LWS)

The variable local window size (LWS) approach is based on the method outlined in Pouliot et al. (2002) and originally adapted from Wulder et al. (2000). In this approach, initially detected local maxima are assessed using locally resized circular sample windows to determine final estimates of local maxima positions. The local window size is determined based on an estimate of spatial dependence. Tree crowns tend to be highly spatially dependent objects in images, therefore local estimates of spatial dependence can produce reasonable estimates of crown size (St-Onge and Cavayas, 1995; Hay et al., 1997; Lévesque and King, 1999; Butson and King, 2004). To derive an approximate measure of the ratio of spatial dependence to crown size, local transects are extended outward from the local maxima being processed. The length of the transects is initially set to that of the largest expected crown in the imagery. Each transect is subsequently reduced using a fourth-order polynomial by removing tail end transect data until a minimum acceptable r^2 value is obtained for the polynomial fit. This acts as a form of scaling to help ensure that transect length is appropriate for the crown being processed. The strongest edge (largest value in the first derivative) within each reduced transect is taken as a measure of the crown radius in that direction, and the mean for all transects is taken as the final crown radius estimate and used as the window size for local maxima detection. In this study, an initial transect sample distance of 15 pixels, with 36 transects separated by 10°, was used for each local maximum. Increasing the number of transects did not improve the results. To scale transects, an r^2 threshold of 0.98 was used, but any value above 0.96 produced very similar results. These parameters were applied for all datasets.

Local smoothing factor (LSF)

The local smoothing factor (LSF) approach was motivated by the success of the LWS methods identified in previous research (Pouliot et al., 2002) but with modifications to address its limited ability to detect local maxima of trees growing in close proximity in relation to the locally determined window size. The approach is essentially the same as the LWS approach but, instead of estimating window size, the amount of local smoothing required to appropriately detect the tree crown is estimated. Estimating the smoothing factor rather than a sample window reduces the probability of omission error because detection is carried out using the minimum window size of 3 × 3, whereas windows can be any size in the LWS approach. As in the LWS approach, initial local maxima are detected and reassessed using local transect analysis to estimate the crown size. The crown size estimate is converted to a Gaussian smoothing factor using an empirically derived equation. A continuous surface is created from the estimated smoothing factors using inverse distance weighting interpolation with a minimum of three points. This surface is then used to locally smooth the image by creating a unique Gaussian filter for each pixel based on the estimated smoothing value. **Figure 3** shows an example image, the surface of estimated σ smoothing factors, and the resulting locally smoothed image. To estimate the final tree locations, the 3 × 3 local maxima detection filter is passed over the locally smoothed image.

The equation used to estimate the smoothing factor from the local transect based crown size estimate was determined empirically using 15 crowns extracted from the imagery for each of the three datasets in the study. Approximately five

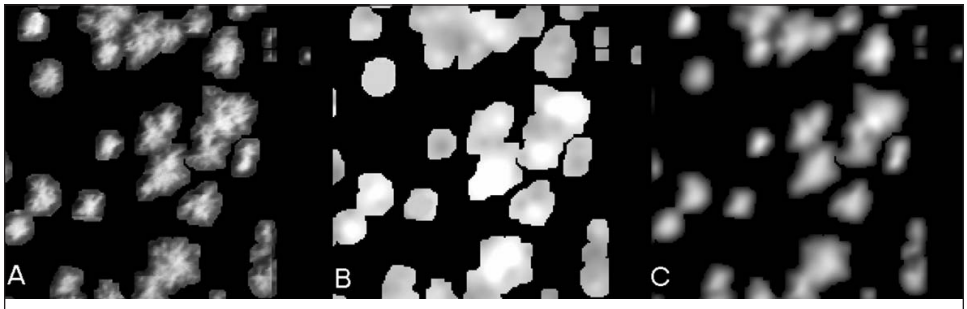

Figure 3. Example of (A) original image, (B) smoothing factor image, and (C) locally smoothed image.

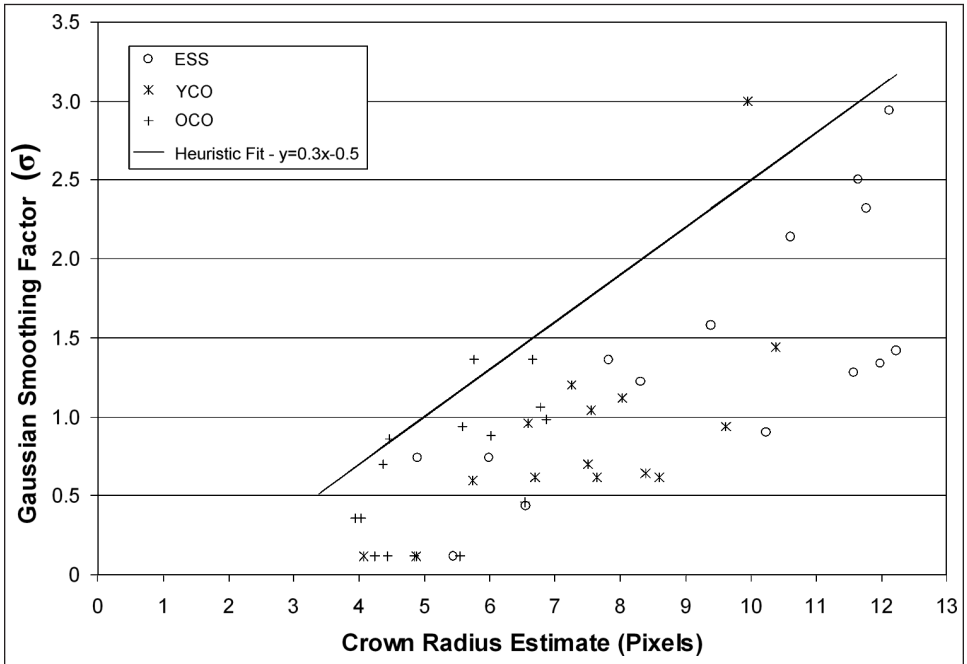

Figure 4. Relation between mean local edge estimate for a crown and degree of Gaussian smoothing required to detect only one local maximum for the crown object.

crowns were selected in each of three size categories of small, medium, and large. The amount of smoothing required to produce only one local maximum for each of these crowns was determined as well as the crown size estimate from the local transect edge detection method. **Figure 4** shows the relation between crown size and Gaussian smoothing factor that was applied to force the crown to contain a single local maximum.

It was expected that this relation would be highly data dependent, requiring separate relations to be defined for each dataset. **Figure 4** reveals, however, that this dependence was not as strong as anticipated and suggests that a generalized equation could be used. Thus, a line was fit to the upper bound of the data for all datasets by modifying the slope and offset of an initial least squares fit to the data ("Heuristic Fit" in **Figure 4**). The equation for this line ($y = 0.3x - 0.5$) was used to determine the local smoothing factor. Apart from this equation, the same parameter values used in the LWS approach were used.

Post-detection merging (PDM)

In the post-detection merging (PDM) approach, the strategy is changed from optimally scaling data for input to automated detection to examining the initial detection results to make local improvements. The watershed segmentation method described in Persson et al. (2002) is used to estimate local maxima positions and the local minima surrounding each maximum, which crudely represent the boundary of the tree crown. The position of the boundary is directly determined by the initial thresholding or classification method used to extract crown pixels. It may not provide consistent crown boundary delineations owing to brightness variations caused by bidirectional reflectance, optical effects, and topography. To aid detection, however, these boundaries can be useful. Referring to **Figure 5**, each boundary pixel common to two crown segments is examined to find the location of maximum brightness along the boundary referred to as the valley brightness (VB). The brightness values for the local maxima positions of the two crown segments being compared are also

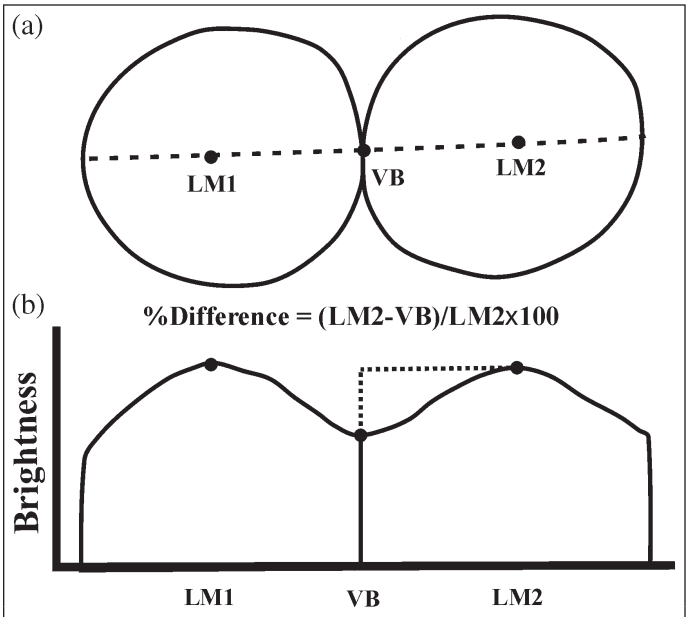

Figure 5. Adjacent crown segments (a) and profile of brightness values (b) from a given local maximum (LM1) through the maximum brightness value along the common boundary (VB) to the adjacent local maximum (LM2). The broken line in the brightness profile identifies the features being compared to determine the merging decision.

extracted (LM1 and LM2). The lower of these two values (LM2) is used to calculate the relative percent difference with the valley brightness. If the percent difference is less than a user-defined threshold, the two crown segments are merged. This process is iteratively applied until no further merging occurs. The selection of the difference threshold is subjective but is the only required parameter in this approach. Tests with various data types and forest conditions could be used to construct a database or relation to provide reasonable estimates for this parameter. In this study, thresholds of 20%, 6%, and 6% were used for the ESS, YCO, and OCO datasets, respectively.

Evaluation of detection results

Detection accuracy was evaluated using a method similar to that of Pitkänen (2001), where points from automated tree detection were assigned to ground-measured reference locations based on an iterative search distance algorithm. In the first iteration, all detected tree points within a given search distance of a reference point were found, the closest was assigned as a match to the reference point, and both points were removed from further consideration. In subsequent iterations, a larger search distance was used up to a maximum search distance of approximately half the size of the average crown size in the dataset. In this study, search distances of 0.1, 0.2, and 0.3 m were used. After the algorithm was complete, reference trees not matched to detected maxima were taken as omission errors, detected trees not matched to a reference tree were taken as commission errors, and matched points were considered correctly identified trees. The numbers of omission, commission, and correctly identified trees were reported along with an accuracy index (AI) that combines errors of omission and commission as shown in the following equation:

$$\mathrm{AI} = [(n - o - c)/n] \times 100 \quad (2)$$

where n is the actual number of trees that exist in the study area, o is the number of omission errors, and c is the number of commission errors. Negative AI values are possible and occur when commission and omission errors are greater than the total number of trees to be detected.

To provide a baseline for comparison of the four optimal detection approaches, the actual optimal global smoothing factor was determined empirically. Each image dataset was smoothed using factors from 1.0 to 3.5 in increments of 0.5 σ, local maxima were detected using the 3 × 3 filtering method, and accuracy results were compiled as described previously. The optimal factor was taken as that which produced the highest AI.

Results and discussion

Empirically derived optimal smoothing

For the empirical baseline analysis, the optimal global smoothing factor for each dataset is shown in **Table 2**. As is consistent with other studies (Wulder et al., 2000; Pouliot et al., 2002), changing the scale by smoothing the data decreased commission error and increased omission error. For all datasets, accuracy initially increased quickly as smoothing increased but slowed as it approached the optimum and then decreased at a slower rate. This trend results from the initial removal of noise by the smoothing process; the noise can be high and causes a substantial amount of commission error. Commission error is further decreased from the reduction of within-crown variability caused by crown branches. In each dataset, commission error due to green background vegetation was not significant in relation to the previously mentioned sources, however, depending on image acquisition timing and the types of vegetation present it can significantly increase commission error. At each successively greater smoothing level, omission error is increased because of small crowns merging with their neighbours or flattening of the crown response. At the smoothing levels highlighted in **Table 2**, the two errors reach an optimal balance. Although the results were compiled for discrete smoothing intervals, they show the approximate optimal smoothing factor to be 3.0, 2.0, and 2.0 for the ESS, YCO, and OCO datasets, respectively, with associated AI values of 87.3%, 66.4%, and 71.8%, respectively. The lower accuracies found for the YCO and OCO datasets are due to high omission error, likely caused by some trees being overtopped by others and hidden from the sensor, small trees with a weak response in the imagery, and multiple trees in close proximity to one another detected as a single tree.

Table 2. Empirically derived tree detection accuracy for different Gaussian smoothing factors, σ, for the ESS, YCO, and OCO datasets.

Scale, σ	Correct	Omission	Commission	Total	Correct (%)	AI (%)
ESS						
1.0	190	7	328	197	96.4	–70.1
1.5	190	7	148	197	96.4	21.3
2.0	190	7	69	197	96.4	61.4
2.5	190	7	29	197	96.4	81.7
3.0	**190**	**7**	**18**	**197**	**96.4**	**87.3**
3.5	190	7	18	197	96.4	87.3
YCO						
1.0	222	43	87	265	83.8	50.9
1.5	210	55	36	265	79.2	65.7
2.0	**201**	**64**	**25**	**265**	**75.8**	**66.4**
2.5	194	71	21	265	73.2	65.3
3.0	190	75	19	265	71.7	64.5
3.5	187	78	19	265	70.6	63.4
OCO						
1.0	118	24	34	144	83.1	59.2
1.5	115	27	20	144	81.0	66.9
2.0	**112**	**30**	**10**	**144**	**78.9**	**71.8**
2.5	101	41	3	144	71.1	69.0
3.0	90	52	4	144	63.4	60.6
3.5	83	59	4	144	58.5	55.5

Note: The optimal factors are highlighted in bold. The correct, omission, commission, and total columns show numbers of trees.

Table 3. Detection error and accuracy for the four optimal detection approaches LMSR, LWS, LSF, and PDM.

Dataset	Correct	Omission	Commission	Total	Correct (%)	AI (%)
LMSR						
ESS	190	7	18	197	96.4	87.3
YCO	201	64	25	265	75.8	66.4
OCO	115	27	16	144	81.0	69.7
Avg.	169	33	19	202	84.4	74.7
LWS						
ESS	188	9	3	197	95.4	93.9
YCO	192	73	18	265	72.5	65.7
OCO	116	26	19	144	81.7	68.3
Avg.	165	36	13	202	83.2	76.0
LSF						
ESS	190	7	41	197	96.4	75.6
YCO	208	57	28	265	78.5	67.9
OCO	115	27	9	144	81.0	74.6
Avg.	171	30	26	202	85.3	72.7
PDM						
ESS	175	22	70	197	88.8	53.3
YCO	210	55	48	265	79.2	61.1
OCO	109	33	14	144	76.8	66.9
Avg.	165	37	44	202	81.6	60.4

Note: The correct, omission, commission, and total columns show numbers of trees.

Optimal detection approaches

The results of the four optimal detection approaches are shown in **Table 3** and for example plots in **Figure 6** for visual evaluation. Comparing these results with the empirically derived optimum reveals that no single approach outperformed the empirical results for all datasets. The differences in the accuracy of the best approaches and the empirical results were small, however, and three of the approaches outperformed the empirical results for specific datasets. The better performance

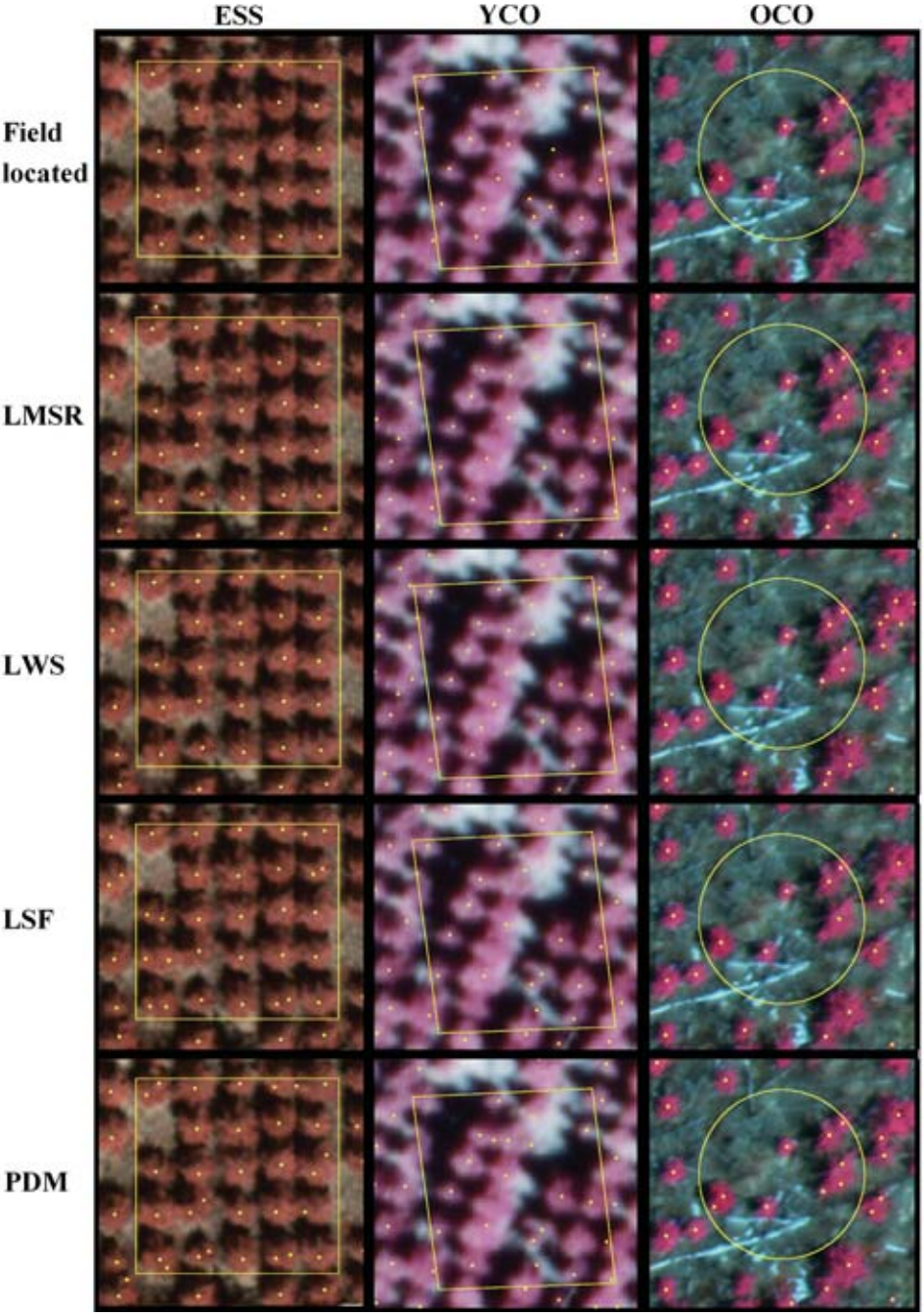

Figure 6. Example detection results for the three datasets (ESS, YCO, OCO), reference tree locations (Field located), and four detection methods compared (LMSR, LWS, LSF, and PDM).

illustrates the potential advantage of local-scale analysis compared to global-scale analysis.

Number of local maxima and smoothing relation (LMSR)

The estimates of the optimal global scale using the LMSR approach are shown in **Figure 7**. For the ESS and YCO datasets, LMSR found the same optimal smoothing level as the empirically determined optimum of 3.0 and 2.0 σ, respectively. In the OCO dataset, however, the optimal smoothing level was underestimated at 1.6. Compared with the empirical results, this difference reduced the AI by only 1%. Of the four approaches, LMSR was the most consistent, as it was second best in the average of each category of error and accuracy over all datasets. It is also the simplest method to implement and the least parameter dependent. Only the range and interval for which the local maxima smoothing relation is desired need to be specified. If incorrectly specified, the results will clearly reveal such errors, as the relation will be difficult, if not

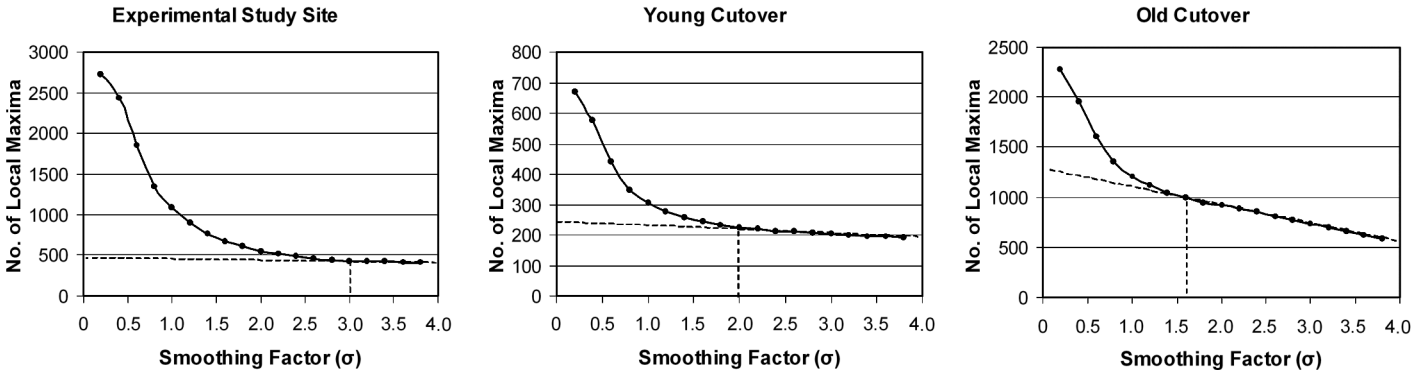

Figure 7. Local maxima smoothing relation for the three datasets used in the study. Broken lines indicate the visual method used to determine the optimal Gaussian smoothing factor.

impossible, to interpret. Processing time may be high for large datasets but was not slower than any of the other methods tested here. A simple solution if processing time becomes problematic is to sample the imagery for analysis.

The main disadvantage of the LMSR approach is that it is subjective and requires users to calibrate themselves to ensure that consistent results are obtained. We had two additional colleagues examine the LMSR curves generated in this study and found that the estimated smoothing factor varied by ±0.2. For the data presented here, this would produce changes in AI of approximately 4%. This approach is also limited by the degree to which commission and omission error can be minimized using a single global scale. For example, consider a dataset with two quite different dominant crown sizes with each crown size being equally represented. The larger crowns may not be detected as single crown objects without applying such a high level of smoothing that many of the small crowns are omitted. Fortunately, such extreme conditions appear to be rare, but users should be aware that the method only seeks to achieve the optimum balance of omission and commission error. In cases where the crown size distribution is skewed, this approach will find the optimal error balance, but less frequent crown sizes will have greater detection error. Clearly, the best results will be obtained when the crown sizes are equally distributed with a narrow standard deviation.

Variable-sized local sample windows (LWS)

The LWS approach produced the highest average accuracy index of all methods but was not the most consistent. It produced very high accuracy for the ESS dataset, which was a highly controlled experiment with the least variable crown sizes and constant tree spacing. These are ideal conditions for this approach because it minimized error caused by large sample windows containing several trees. This method performed relatively poorly in the two operational regeneration datasets with inconsistent spacing and more variable crown sizes. It also required specification of the local transect sample distance, the number of transects, and the r^2 scaling factor. Of these, the most critical is the sample distance specification, as too large a sample distance may extend over more than one crown, even after transect scaling. This can occur when a small crown is next to a larger one, in which case the polynomial achieves a sufficient fit with the larger crown because the small crown is treated as noise. Thus, this approach appears to be ideally suited for even-aged plantation forests with equally sized and spaced crowns, as a constant sample distance can be specified that is appropriate to all crowns in the imagery. Further, the constant tree spacing greatly reduces the problem of large window sizes containing more than one tree crown.

Local smoothing factor (LSF)

The LSF approach performed the best of all methods with the datasets that had highly variable crown sizes and tree spacing (YCO and OCO). For the OCO dataset, which had the highest crown size variability, the observed AI was ~5% higher than the next best optimal detection approach. This is an important characteristic because these datasets are representative of more operational conditions. Further, it shows that local approaches can achieve higher accuracy than global approaches. However, the improvement comes at the cost of an increased number of parameters and parameter sensitivity. Such parameter dependence strongly limits the utility of an approach, as it becomes more difficult to understand the nature and sources of errors. The poor results obtained for the ESS dataset suggest that the heuristic equation defined to convert the crown size estimates to smoothing factors was inappropriate for this dataset. To better evaluate the effect of the equation on the results, the slope of the original equation was increased to 0.35. Using this slope, an AI of 80.3% was obtained, representing an increase of about 5%. This clearly illustrates that this approach is very sensitive to the user-defined equation and it is unlikely that it would be useful for operational applications. Further research may identify improved methods to determine this equation in a more consistent and objective manner than that used here. Although this approach had several limitations, an approach designed to achieve the same objectives, i.e., local scale optimization, is needed to provide the highest possible accuracy using automated tree detection in high-resolution imagery. Currently, no method has been identified that can accomplish this task.

Post-detection merging (PDM)

The PDM approach produced the lowest accuracies of all methods tested. This is due in part to the global threshold used for the merging decision, which is unable to account for local variations in brightness that produce variations in the depth of brightness valleys between crowns in the imagery. Further, using the maximum value of the common boundary between two crowns often led to incorrect merging decisions. This occurred when a small portion of the shared boundary contained brightness values similar to the local maxima of the crown segments being compared, but the majority of the boundary distinctly divided the two segments and suggested that they should not be merged. Using the average brightness value of the common boundary produced poorer results because the average was biased by the darker edges of the crown, making the boundary appear stronger than it actually was and leading to numerous merges being missed. Owing to these limitations, this approach is not useful as an error-minimization approach on its own but could be useful as a low-priority check to merge trees that strongly meet defined merge criteria.

Comparison with other research

It is difficult to compare the results of this research with those from other studies because most were conducted in mature forests, and sensor types and initial image resolutions were not the same as those used here. With regard to sensor and image characteristics, two of the most critical are spatial resolution and sun–sensor geometry. The spatial resolution dictates the number of omission errors, as some crowns may be too small to be detected at larger resolutions and other crowns that are closely spaced may be merged into a single crown. The methods presented in this study and reviewed here require very high spatial resolution imagery in relation to the crown size. If large area coverage is required, however, coarser resolution must usually be used with an associated reduction in individual tree detection accuracy. For example, Gougeon and Leckie (1999) have taken this approach to emulate more operational conditions for automated detection in regenerating forests. Their results were promising under the design criteria but, as expected, numerous omission errors were observed. The sun–sensor geometry is another important factor because it affects both the radiometric and geometric properties of the tree crowns. Using simulated imagery of mature eucalyptus forests, Culvenor (2000) showed that tree detection accuracy was best with small off-nadir view angles, with backscattered as opposed to forward-scattered radiance, and at higher solar zenith angles.

Forest properties control how visible and distinct crowns are in images. Coniferous trees are generally much easier to accurately detect than deciduous species due their distinct radiometric crown morphology (Warner et al., 1999). The results presented in this study showed the effects of varying tree size, density, and spatial pattern, with different detection approaches identified as optimal depending on these conditions. A distinct difference between mature and regenerating coniferous forests is that mature forests typically have reduced occurrence of trees in close proximity because of self-thinning caused by interspecific and intraspecific competition. The greater spacing between trees will generally enhance the distinct nature of individual crowns, improving detection results. However, mature forests tend to have greater vertical structure than regenerating conditions, and trees in the lower canopy positions can be difficult to detect if hidden from the sensor or strongly shaded by other crowns. This can also occur in regenerating forests, where smaller trees are hidden by larger ones, but it is less frequent.

Variations in accuracy evaluations also render comparisons among studies difficult. In regeneration conditions, it is a simpler task to create field maps of tree locations. In mature forests, greater error in locating individual trees is likely due to errors associated with classical or global positioning system (GPS) survey techniques. Accuracy evaluations vary with regard to the types of comparisons made, including plot based versus single tree based, manually interpreted versus field data, and the selection of data to use as reference. For example, Gougeon (1995) compared automated results with ground counts on a plot basis and on an individual tree basis with manual interpretation. Erikson (2003) did not include commission error occurring in areas considered as noncrown background. Validation by manual interpretation also tends to produce higher accuracy measures than field-based validation because it does not include errors occurring from trees not visible to the sensor or from two trees in close proximity that appear as a single crown. For example, AI calculated from the results of Brandtberg and Walter (1998) was 8% less using field reference data than using reference data from image interpretation.

Despite a lack of capability for rigorous comparison among studies, it is still of interest to review previous research results and theoretical aspects in light of this study. In the following, where possible, the accuracy values published in other studies have been converted to the AI (Equation (2)) used in this study for comparison purposes. One approach that was not considered for this study is scale space analysis. It has been well developed and is theoretically well justified in computer-vision research as a means of extracting basic image structures over multiple scales. Thus, it appears to be well suited for automated tree detection. However, current scale space approaches have not shown significantly improved accuracies over other methods and are often dependent on the definition of many difficult to define parameters. The scale space edge detection method proposed by Brandtberg and Walter (1998) showed detection AI to be 59% in mixed, mature European forests. The method required several input parameters–thresholds, and it was noted that it had difficulty detecting crowns that did not have rounded or circular shapes. In another study using lidar data of leaf-off mature forests in the eastern United States, Brandtberg et al. (2003) determined suitable parameters for a scale space approach based on blob detection using field data. Other more simple approaches have produced

reasonable results and are equally or less parameter dependent. The variable window sized approach tested in mature natural and plantation forests by Wulder et al. (2000) produced an AI of 51%. In regeneration conditions, Pouliot et al. (2002) found an AI of 89% using the same approach with the ESS dataset as that used in this study. The slightly higher results reported in this paper are due to improvements in the extraction of crown pixels, using classification rather than simple thresholding. Although the variable window size approach has been successful, as shown in this study, it is not appropriate for forests with highly variable crown sizes and spacing. Local binarization has been tested as a means to improve results of local maxima crown detection. In mature European forests of varying species compositions, Pitkänen (2001) observed AI values ranging from 48% to 96% depending on parameter values used and specific forest conditions. Improvements with local binarization ranged from 0% to 4%. A local surface fitting method was used by Persson et al. (2002) to determine which of three scales was most appropriate for automated tree detection. The approach worked well, with lidar data of Norway spruce forests producing an AI of 70%. The main disadvantage of the approach is that the scales were discrete and selected arbitrarily. Further, it is not clear what the advantages and disadvantages of lidar data were compared with optical data for tree detection using this method. Leckie et al. (2003) found the two data sources to be complementary, each having its own unique advantages. In their study, lidar data proved to be better suited for more open forests and optical data better in moderate to closed canopy conditions. Erikson (2003) used a region-growing approach, where the size of the crown to be segmented was estimated using the largest curved segment from a cluster determined by image binarization. Reported accuracy for the mature European forest of mixed composition was AI ≈ 81% with manually interpreted delineations. As with many region-growing algorithms, the method requires numerous parameter specifications to achieve accurate results.

The results of this study show that the approaches tested produced accuracies within the range of those reported in other studies. Further, as shown in other studies, the accuracy obtained was dependent on the specific forest conditions and number of parameters used. Accuracies can be high for methods that utilize numerous well-estimated parameters. Other more general methods produce lower accuracies but are less parameter dependent, making them more widely applicable to a range of forest conditions. Thus, the choice of the optimal method depends on the required accuracy, acceptable cost, and forest conditions.

Conclusion

This paper has presented a direct comparison of different optimization approaches for automated tree detection within a range of datasets representing a variety of sensor and scene characteristics of regenerating forests of varying structure. Of the approaches tested, determination of the optimal global smoothing (LMSR approach) factor produced the most consistent accuracies, was typically second highest in accuracy, and was the least parameter dependent. Its major limitation was that it requires a moderately subjective assessment of the relation between the number of local maxima and smoothing intensity, which requires user calibration to obtain consistent results. Determining an optimal local window size (LWS approach) worked extremely well with the dataset consisting of constant tree spacing and crown sizes, and determining the optimal local smoothing factor (LSF approach) produced the highest accuracies in the more variable tree size and spacing datasets. However, both of these approaches were dependent on appropriately defined input parameters. The post-detection merging (PDM) approach did not perform sufficiently well to merit further consideration.

Image scale is a critical component in automated tree detection, whether it is represented by the smoothing level, window size, or another form of size parameter. Using a global scale representation such as a constant window size, smoothing factor, or threshold can only lead to a certain level of tree detection accuracy. Local methods have the potential to achieve theoretically perfect accuracy but have proven technically difficult to implement. The two local approaches tested here were encouraging, but their reliance on parameters is considered too limiting for more operational applications. Further research will focus on identifying additional approaches for optimal detection and combining some of the more promising ones into an integrated approach.

Acknowledgements

This research was funded by a grant to D. King and a scholarship to D. Pouliot from the Natural Sciences and Engineering Research Council of Canada (NSERC). Image data for the ESS and OCO sites were collected by the Ontario Ministry of Natural Resources Data Acquisitions Branch. Field data for these sites were collected by the Ontario Ministry of Natural Resources. Image data for the YCO were acquired by Wiskair Ltd., and Duncantech Inc. (now Redlake Inc.) generously loaned the MS3100 digital camera. Field data collection for the YCO dataset was greatly aided by the assistance Mark Lindsay. Buchanan Forest Products Ltd. provided historical site information and logistic support for image and field data collection of the YCO dataset. Thanks to Wayne Bell of the Ontario Forest Research Institute and Doug Pitt of the Canadian Forest Service for advice and support in this research.

References

Bell, F.W., Ter-Mikaelian, M.T., and Wagner, R.G. 2000. Relative competitiveness of nine early-successional boreal forest species associated with planted jack pine and black spruce seedlings. *Canadian Journal of Forest Research*, Vol. 30, No. 5, pp. 790–800.

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

Brandtberg, T., Warner, T.A., Landenberger, R.E., and McGraw, J.B. 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. *Remote Sensing of Environment*, Vol. 85, No. 3, pp. 290–303.

Butson, C.R., and King, D.J. 2004. Determination of optimal forest sample plot spatial extent using lacunarity analysis of airborne imagery. In *Proceedings of the 18th Biennial Workshop on Color Photography and Videography in Resource Assessment*, 16–18 May 2001, Amherst, Mass. American Society for Photogrammetry and Remote Sensing, Bethesda, Md. CD-ROM.

Culvenor, D.S. 2000. *Development of a tree delineation algorithm for application to high spatial resolution digital imagery of Australian native forest*. Ph.D. thesis, University of Melbourne. Melbourne, Australia.

Dralle, K., and Rudemo, M. 1996. Stem number estimation by kernel smoothing in aerial photos. *Canadian Journal of Forest Research*, Vol. 26, pp. 1228–1236.

Erikson, M. 2003. Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, Vol. 33, No. 8, pp. 1557–1563

Gougeon, F.A. 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution images. *Canadian Journal of Remote Sensing*, Vol. 21, No. 3, pp. 274–288.

Gougeon, F.A., and Leckie, D.G. 1999. Forest regeneration: individual tree crown detection techniques for density and stocking assessments. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, 10–12 February 1998, Victoria, B.C. Edited by D. Hill and D. Leckie. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, B.C. pp. 169–177.

Hay, G.J., Niemann, K.O., and Goodenough, D.G. 1997. Spatial thresholds, image-objects, and upscaling: a multiscale evaluation. *Remote Sensing of Environment*, Vol. 62, No. 1, pp. 1–19.

Hay, G.J., Blaschke, T., Marceau, D.J., and Bouchard, A. 2003. A comparison of three image-object methods for multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 57, pp. 327–345.

Lautenschlager, R.A., Bell, F.W., Wagner, R.G., and Winters, J.A. 1997. The Fallingsnow Ecosystem Project: comparing conifer release alternatives in northwestern Ontario. *Forestry Chronicle*, Vol. 73, No. 1, pp. 35–38.

Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., and Shreenan, R. 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. *Canadian Journal of Remote Sensing*, Vol. 29, No. 5, pp. 633–649.

Lévesque, J., and King, D.J. 1999. Airborne digital camera image semivariance for evaluation of forest structural damage at an acid mine site. *Remote Sensing of Environment*, Vol. 68, No. 2, pp. 112–124.

Lindeberg, T. 1994. *Scale-space theory in computer vision*. Kluwer Academic Publishing, Dordrecht, The Netherlands.

Persson, A., Holmgren, J., and Soderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, Vol. 68, No. 9, pp. 925–932.

Pinz, A. 1999. Tree isolation and species classification. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, 10–12 February 1998, Victoria, B.C. Edited by D. Hill and D. Leckie. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, B.C. pp. 127–139.

Pitkänen, J. 2001. Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. *Canadian Journal of Forest Research*, Vol. 31, No. 5, pp. 832–844.

Pitt, D.G., Wagner, R.G., Hall, R.J., King, D.J., Leckie, D.G., and Runesson, U. 1997. Use of remote sensing for forest vegetation management: a problem analysis. *Forestry Chronicle*, Vol. 73, No. 40, pp. 459–478.

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*, Vol. 82, Nos. 2–3, pp. 322–334.

St-Onge, B.A., and Cavayas, F. 1995. Estimating forest stand structure from high resolution imagery using the directional variogram. *International Journal of Remote Sensing*, Vol. 16, No. 11, pp. 1999–2021.

Walsworth, N.A., and King, D.J. 1999. Image modeling of forest changes associated with acid mine drainage. *Computers and Geosciences*, Vol. 25, pp. 567–580.

Warner, T.A., Lee, J.Y., and McGraw, J.B. 1999. Delineation and identification of individual trees in the eastern deciduous forest. In *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, 10–12 February 1998, Victoria, B.C. Edited by D. Hill and D. Leckie. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, B.C. pp. 81–91.

Wulder, M., Niemann, K.O., and Goodenough, D.G. 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment*, Vol. 73, No. 1, pp. 103–114.