Detecting individual eucalyptus crowns in aerial photographs using template matching and classification

Maciel Zortea[1]
Marcelo Nery dos Santos[1]
Bianca Zadrozny[1]
Emerson Roberto Schoeninger[2]
Cristiano Cardoso Stetz[2]

[1] IBM Research Brazil
Avenida Pasteur 138 – 22290-240 – Rio de Janeiro – RJ, Brazil
{mazortea,mnerys,biancaz}@br.ibm.com

[2] Klabin
Avenida Brasil 26 – 84275-000 – Telêmaco Borba – PR, Brasil
{eschoeninger,ccstetz}@klabin.com.br

Abstract. Collecting and analyzing forest information is critical for sustainable forest management. Here we propose a method to automatically detect the location and diameter of tree crowns at early growth stages in regularly planted forests using very high spatial resolution RGB imagery acquired by unmanned aerial vehicles (UAVs). A list of candidate detections is generated matching the multiscale convolutions of synthetic crown templates to image objects. Strong matches are filtered using color-based rules. Local attributes describing the color and spatial information in small image patches centered in each retained detection are passed to an off-line trained Random Forest classifier that assigns a level of confidence to each tree crown detection. The method is tested on orthorectified RGB mosaics with a pixel spacing of about 11 cm using circular templates with diameters in the range 50–200 cm. Experiments at two study sites containing about 120-day-old plantations of eucalyptus, located in Southern Brazil, suggest detections accuracies above 90% when non-overlapping adjacent crowns have a diameter larger than 6 pixels and are surrounded by mixed backgrounds such as exposed soil and debris from the previous harvest. The automated counts of trees in 12 footprints of 1,257m^2 were within $\pm$11% of the visual estimate, and within $\pm$4% when averaged for the study. Examples of challenging scenarios requiring further methodological developments are presented. We anticipate that automated tree crown detection using the proposed prototype algorithm may complement traditional field-based tree inventory.

Keywords: forest inventory, image segmentation, remote sensing, inventário florestal, segmentação de imagens, sensoriamento remoto.

1. Introduction

Forests are an asset for economic development and for the planet. The sustainable management of planted forests of pine and eucalyptus requires the monitoring and analysis of large volumes of data. To meeting the demand for raw materials, a precise estimate of the number of trees growing in the stands is highly desirable. Counting trees is part of the costly forest inventory. According to Oliveira et al. (2014), the continuous forest inventory is one of the most widely used inventory forms in Brazil to monitor forest growth. However, errors associated with conventional inventories have been reported, such as bias in the measurement of tree diameter and height, plot area, data manipulation and, mainly, errors associated with the sampling procedure. The impact of such errors vary with the dendrometric variable analyzed and are higher for the number of trees per hectare. These errors decrease the quality of the data collected and negatively affect the accuracy of the growth and production models, impacting planning. Such errors can only be solved if the census of trees is carried out in the whole area. There is a great variation in tree spacing in planted forests, which can be caused by errors in the planting process, the presence of topographic conditions that prevent the planting operation correctly, and natural mortality in the forest, among others. All these issues impact the estimated number of trees and, consequently, the total volume of wood. Remote sensing can complement field-based assessments. For instance, UAVs offer flexibility in acquiring images, at lower costs and improved temporal and spatial resolution if compared to satellites. Poor flight stability, limited autonomy, and camera calibration issues are among the challenges to overcome.

1.1. Objectives

The objectives of this research are two-fold. First, to develop and test a prototype of an automated algorithm for individual tree crown detection using well-established image processing and statistical learning techniques. The method should be able to detect planted trees at early stages of development analyzing georeferenced very high spatial resolution ($\approx$10 cm) RGB mosaics of aerial photographs. Secondly, to understand practical challenges that motivate the need for the development of more advanced image analysis and machine learning methods tailored for this specific problem.

1.2. Related work

Ke e Quackenbush (2011) grouped previous individual tree crown detection studies using passive remote sensing data in four broad algorithmic categories: based on local maximum filtering, image binarization, scale analysis, and template matching. These methods proved useful to analyze forests in a wide range of age and use conditions. Pouliot et al. (2002) presented a tree detection and delineating algorithm that was tested on airborne imagery of 5 to 15 cm pixel spacing to detect 6-years-old planted trees based on the analysis of transects extending outward from a potential tree apex. Kang et al. (2016) relied on near-infrared UAV imagery to identify tree crowns in undulating areas. The study focuses on evaluating different methods performance with focus on sunlit and shaded areas, as well as, sparsely and densely populated forests. LIDAR could point data helps crown detections (Oliveira et al. (2014)).

The remaining of the paper is organized as follows. Section 2 describes the proposed prototype algorithm and the main contributions of this work, which include the definition of a customized template for crown detection, and the use of simple image patch descriptors to assign levels of confidence to the automated detections. The accuracy of our algorithm is measured on image footprints randomly selected from two real data sets presented in Section 3. Conclusions are summarized in Section 4.

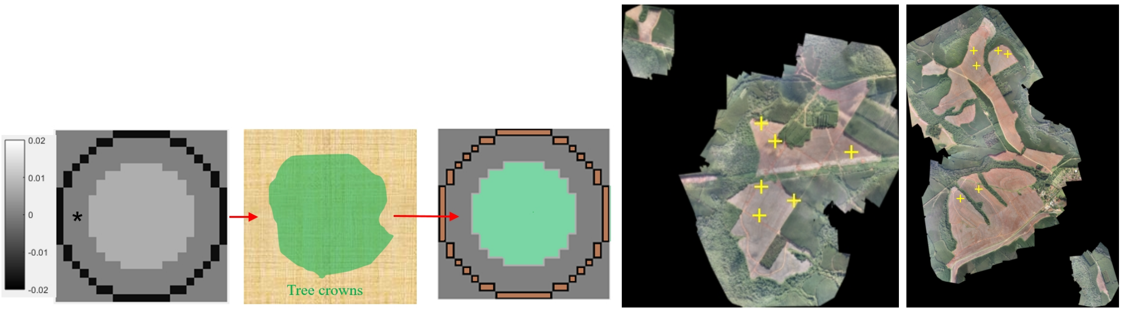
2. Methodology

Our tree crown detection pipeline involves the five main steps presented below.

2.1. Pre-processing: adaptive contrast enhancement

The colors of the input image are enhanced using a locally adaptive contrast enhancement approach. We compute the range of intensity values in a fixed neighborhood of radius r (r=10m in our experiments) around each pixel, for each color channel. This estimates the approximated color of the darker objects (usually trees or shadows) and the brighter image background typically present in our case studies. The intensity of each channel in the local neighborhood is then linearly mapped to new values such that the resulting image stretches to 8 bits. Morphological operations such as the “top-hat"filter (Serra e Soille (2012)) helps in the

ISBN: 978-85-17-00088-1



(a) (b) Data set 106 (c) Data set 108

Figura 1: (a) Proposed synthetic template combining a disk and ring elements used for multiscale convolutions. (b–c) Overview of the two test sets showing the location of the center of the 12 randomly selected image footprints used for accuracy assessment (as a scale reference, all crosses are 150 m wide).

implementation. A Gaussian low-pass filter with a kernel size σ (σ=0.125 m in our experiments) helps to remove noise in the resulting image.

The primary objective of this preprocessing is not to provide a visually enhanced image preserving radiometric information since color distortions (tone) are introduced in the process of analyzing independently each input channel. This preprocessing step is only intended to attenuate eventual uneven illumination effects along large input mosaics (due to variations in photo exposure, cloud shadows, etc), and assist the successive image processing tasks, as it may be a bit easier to set, for instance, color channel thresholds.

2.2. Compute band indices to help crown detection

We would like to have discriminating images where tree crowns, shadows, and other land covers are well distinguishable. However, achieving a good statistical separation is in general difficult using solely RGB imagery without infrared bands. Given the pre-processed image, we compute the auxiliary band index:

$$\rho_G = 2\frac{G}{(R+B)} \tag{1}$$

i.e., we measure the excess of the green color component (G) compared to the average of the red (R) and blue (B) values. Observation suggests that often the planted tree crowns have a higher index (Eq. 1) then the surrounding background areas in the image. Therefore, thresholding the resulting ρ_G image is expected to help to separate tree crowns from other spurious candidate detections. Furthermore, we retain the difference (grayscale) image $\delta = (G - R)$ for template matching. Strong candidates detections obtained convolving synthetic tree crown templates to image objects, and above the color index threshold t_{ρ_G}, will be retained for final classification.

2.3. Mutiscale convolutions

Convolving an image with an appropriate filter is a popular approach to emphasize objects or structures of interest in digital images (Trier, Zortea e Tonning (2015)). A relevant question is how to set the appropriate template for a specific application. In the tree crown detection problem, possible attempts include deriving a shape based on data observations, for instance averaging a set of representative tree crowns, or using synthetic circular templates. Here we focus on the second approach, proposing the use of circular templates made of an internal circle, aimed to cover most of the tree crown area, and an external ring that is expected to lie on the adjacent image background (e.g. exposed soil, debris, etc) as depicted in the sketch shown

ISBN: 978-85-17-00088-1

in Figure 1(a)[1]. We keep a fixed 50 cm margin between the inner circle and the outer rim to accommodate tree crown with irregular shapes.

Once defined the template, the grayscale difference image δ is convolved with the template shown in Figure 1(a) at increasing spatial sizes to detect tree crowns of multiple diameters. A relevant aspect is choosing the right scale of the templates. Ideally, the range of suitable scales should be based on a priori information about the typical sizes of the trees to be detected in a study area, and the spacing between adjacent trees. For each pixel location in the resulting convolution image I_c, we store the maximum convolution value and the respective (optimal) scale fitting the image objects. In the sequence, we mark local maxima in the convolution image I_c inside a circular neighborhood of radius r_m (r_m=75 cm in our experiments), and retain the spatial locations of the strong maxima, i.e., convolution values above a certain threshold t_c (set at the 70-th percentile of I_c in our experiments). The resulting candidates having a color value above $t_{\rho G}$ in Eq. 1 (we use a loose $t_{\rho G}$=0.85 in our experiments) and below a fixed red color value t_r (set at the median of R in our experiments) are retained as candidates for final classification and confidence estimation.

2.4. Feature extraction for classification and confidence estimation

For each candidate point detection retained after the above-mentioned steps, the convolution response and color values in the local image patch centered in each candidate are further analyzed. Let $I(i,j)$ be the small image patch centered at pixel locations (i,j) of size $w = 1.5\phi$, where ϕ is scale of the best fitting circular template matched to the image (object), which correspond to the estimated diameter of the tree crown, rescaled to a fixed $w \times w$ matrix (w=9 in our experiments). Each $I(i,j)$ is first centered in its median value and then converted to the 18-dimensional feature vector:

$$F(I) = [u_1, u_2, \ldots, u_9, v_1, v_2, \ldots, v_9], \tag{2}$$

where the sum of rows and columns of the image patch:

$$u_k = \sum_{j=1}^{9} I(k,j) \text{ and } v_k = \sum_{i=1}^{9} I(i,k), \tag{3}$$

help to describe each patch. Each candidate detection will be represented by the concatenated 39-dimensional features $x = \{F(R), F(I_c), \sigma^2_{(R)}, \sigma^2_{(I_c)}, \sigma^2_{(R,I_c)}\}$ extracted from the patches of the red channel R and the convolution image I_c. Here $\sigma^2_{(\cdot)}$ are the variances and the covariance of the pixels in $I(i,j)$ for the channels R and I_c. These descriptors are straightforward to compute. Possible alternatives include using more sophisticated keypoint descriptors and textures.

2.5. Statistical classification

We estimate the probability that each candidate detection x is a tree crown using the Random Forest classifier (Breiman (2001)). This popular non-parametric model has proved useful in many machine learning applications and was selected because its posterior estimates were often not saturated in 0 or 1 making it a bit easier to group classification probabilities in a few wide intervals. This binary classifier was trained off-line using a small labeled set of tree and background detections. Out-of-bagging classification error was below 2%. The trained model is kept unchanged, i.e., we do not retrain the classifier for each new image to the analyzed. Detections with a very low posterior estimate (confidence) are discarded (P(tree|detection) <

[1] For convenience, the sum of all pixels values inside the template are set to zero.

0.05). The coordinates of the remaining detections, the estimated crown diameters, and the respective probabilities constitute the output to the user. To help visual interpretation, the probabilities are grouped in a few discrete intervals that are color coded to overlay the detected crowns to the input image (see examples in Figures 2–3).

3. Experiments

3.1. Data

The proposed algorithm was tested on 12 circular footprints of diameter 40 m (1,257 m^2 each) randomly sampled from two large image mosaics named 106 (size 760 MP) and 108 (size 1.59 GP), shown in Figure 1(b–c). The sites are located in Telêmaco Borba, Paraná State, Southern Brazil, near (24.2363S, 50.4433W) and (24.2189S, 50.5532W). The study sites are commercially used for pulp production. The sampled areas contain about 120-day-old eucalyptus planted at the nominal spacings of 3.80×1.57 m (site 106) and 3.55×1.69 m (site 108). Six locations were sampled from each of the two data sets. The aerial photographs were collected using the Carcará UAV by a commercial partner (SantosLab Ltda, Rio de Janeiro, Brazil) and processed to orthorectified mosaics using the PiX4D© software (Pix4D SA, Lausanne, Switzerland). The output imagery was stored in GeoTIFF files georeferenced in UTM coordinates with spatial resolutions 10.6 cm and 11.5 cm, respectively.

3.2. Protocol

The random sampling was guided by a shapefile delineating the approximate location of the planted fields to be processed by the proposed algorithm. Each footprint had to contain more than 50% of its area planted to be eligible for accuracy assessment, and be nearly cloud-free. The alive (green) tree crowns in each image footprint shown in Figures 2–3 were visually marked overlaying a 10 pixel reference disk centered at each tree crown location. The diameter of the footprint was set to tradeoff the laborious image interpretation work while sampling over 100 trees per footprint. Because of the small tree diameters (sometimes smaller than 5 pixels), and the resulting color of mixed pixels, there were challenging scenarios to establishing visually wheatear tree crowns were present and/or alive.

The binary Random Forest classification model was trained using samples from a spatially disjoint scene that did not overlap the testing areas. We assigned blue color for crowns detected with high confidence (here defined as having posterior probabilities for the tree crown class in the range 0.6–1.0), cyan for medium confidence (0.4–0.6), yellow for low confidence (0.1–0.4), and red color as a baseline for very low confidence (0.05–0.1). These intervals were set by visual inspection of a separate training set.

The proposed tree crown detection algorithm was tested using the proposed circular template with internal circular diameters of $\{0.5, 0.6, \ldots, 2.0\}$ m. Despite not expecting to find such large tree crowns in early stage plantations, we intentionally set the large upper limits to check wheatear (wrong) large detections would be included. We count the number of trees visually in each footprint and compare it with the automated detections provided by the tested algorithm. This is done by checking whether the proposed crown center hits the ground truth circular mask 10 pixels wide manually centered in each tree crown location.

3.3. Results

Figures 2–3 show tree crown detected using the proposed method. The respective accuracy scores are summarized in Tables 1–2. Detections whose centers hit the reference 10-pixel wide disk centered in each tree crown were considered correct. The accuracy of the estimated crown diameter is not evaluated quantitatively, but visual inspection of Figures 2–3 suggest good or

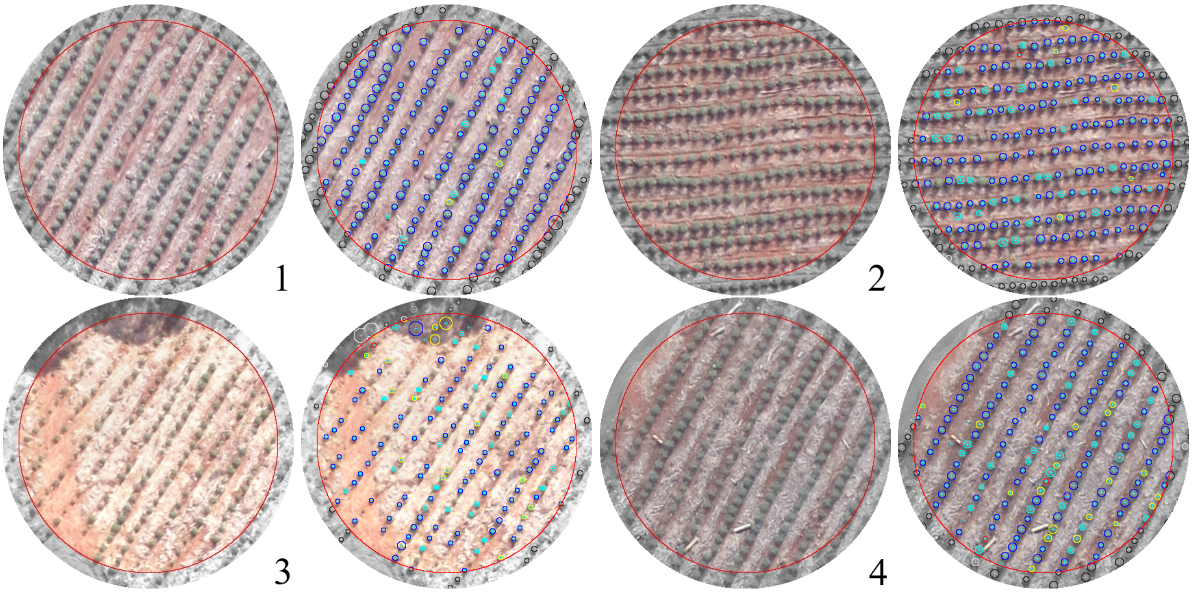


Figura 2: Four of the six footprints located at the study site 106. The red circle of 40 m diameter delineates the area used for accuracy assessment. The image extends to an additional 2.5 m wide external rim area, depicted in grayscale, providing visual continuity and scale reference. Besides each footprint lies the respective input RGB image with the automated detections overlaid (here shown without preprocessing). Circles in blue color correspond to the crowns detected with high confidence, cyan for medium confidence, yellow for low, and red for very low confidence. The five-pixel wide crosses (53 cm) inside the circles indicate the estimated center of the crowns and are shown in cyan in all cases. Individual crosses not surrounded by a circle indicate possible tree locations with confidence below 5% that were excluded by the classification step. Scattered yellow dots show points of local convolution maxima that were excluded earlier during color analysis.

reasonable agreement in most of the footprints tested.

Detections with higher levels of confidences were confirmed to be more likely to be trees. For instance, 94.3% of the 857 detections labeled with high confidence (in blue color in Figure 2) hit the ground truth crown masks according to the adopted criteria. This figure decreases to 87.2% for the 141 detections labeled with medium confidence (Table 1). Although hitting the manually drawn reference crown masks would be reassuring, visual inspection of

Tabela 1: Summary of individual tree crown detections for the six footprints located in the data set 106. The number of detections, (the percentage of detections that hit the reference tree crowns), and the percentage of accumulated detections if compared to the visual counting (100%) are shown for each level of confidence estimated using the proposed algorithm.

#	Visual counting	Level of confidence for the automated detections							
		high	%	medium	%	low	%	very low	%
1	169	148 (96.6)	87.6	9 (77.8)	92.9	2 (100.0)	94.1	0 (0.0)	94.1
2	194	147 (92.5)	75.8	34 (94.1)	93.3	6 (0.0)	96.4	1 (0.0)	96.9
3	148	93 (95.7)	62.8	28 (85.7)	81.8	16 (68.8)	92.6	1 (0.0)	93.2
4	144	104 (95.2)	72.2	28 (92.9)	91.7	18 (72.2)	104.2	2 (100.0)	105.6
5	210	184 (90.2)	87.6	19 (94.7)	96.7	6 (83.3)	99.5	1 (100.0)	100.0
6	194	181 (96.7)	93.3	23 (69.6)	105.2	6 (50.0)	108.2	0 (0.0)	108.2
all	1059	857 (94.3)	80.9	141 (87.2)	94.2	54 (63.0)	99.3	5 (60.0)	99.8

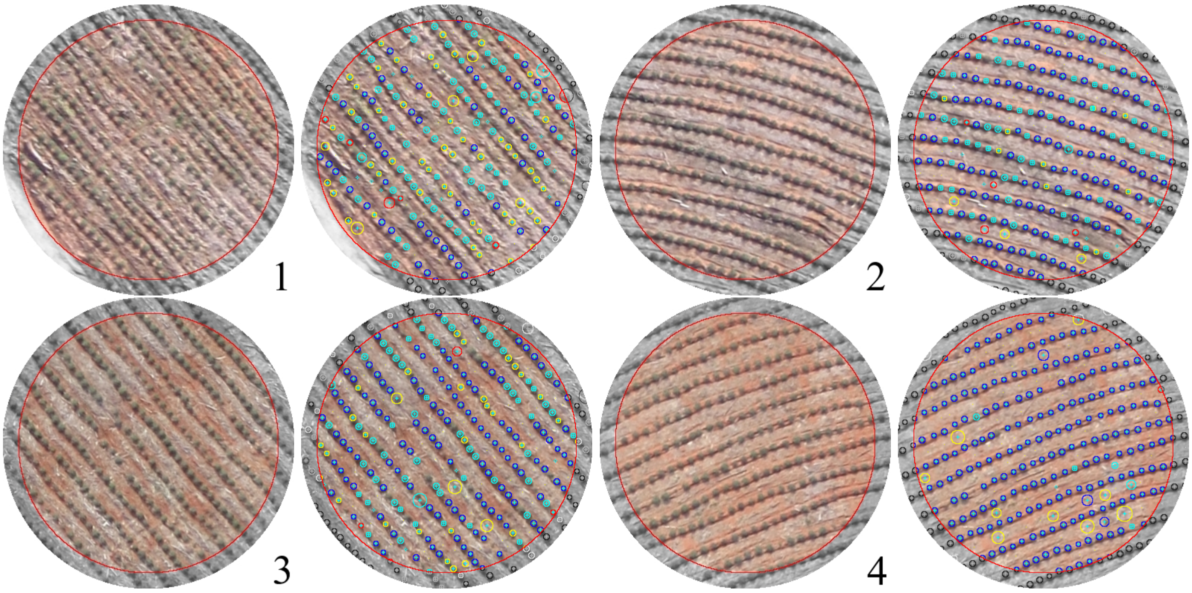


Figura 3: Four footprints located at the study site 108. For reference the cyan crosses are five pixels wide (58 cm). (see caption to Figure 2 for further details).

Tabela 2: Summary of individual tree crown detections for the six footprints located in the data set 108. The number of detections, (the percentage of detections that hit the reference tree crowns), and the percentage of accumulated detections if compared to the visual counting (100%) are shown for each level of confidence estimated using the proposed algorithm.

#	Visual counting	Level of confidence for the automated detections high	%	medium	%	low	%	very low	%
1	184	54 (96.3)	29.3	73 (84.9)	69.0	54 (72.2)	98.4	5 (80.0)	101.1
2	208	135 (95.6)	64.9	71 (97.2)	99.0	12 (75.0)	104.8	4 (25.0)	106.7
3	222	121 (99.2)	54.5	60 (95.0)	81.5	30 (76.7)	95.0	3 (66.7)	96.4
4	204	211 (94.3)	103.4	5 (60.0)	105.9	9 (0.0)	110.3	1 (0.0)	110.8
5	218	217 (97.7)	99.5	5 (80.0)	101.8	6 (0.0)	104.6	2 (0.0)	105.5
6	184	128 (97.7)	69.6	39 (89.7)	90.8	19 (84.2)	101.1	1 (0.0)	101.6
all	1220	866 (96.7)	71.0	253 (90.9)	91.7	130 (66.9)	102.4	16 (43.8)	103.7

Figures 2–3 suggests it may be not strictly necessary in order to accurately estimate the total number of trees. This is because some detections are still close enough to the crown locations if compared to the distance between adjacent trees. The high confidence category included 80.9% and 71% of the reference number of trees for data sets 106 and 108, respectively.

The estimated number of trees was close to the reference established visually when the automated detections allocated in the high, medium, and low levels of confidence were combined. In this setting our model was strongly predictive, counting 99.3% (Table 1) and 102.4% (Table 2) of the trees, which means that the aggregated estimates of the number of trees in each set of six footprints were very close to the reference number established visually. This is because missing detections are partially compensated by wrongly-placed crown detections. Finally, it is important to keep in mind that variations among individual footprints counts do exist and are not negligible, up to -7.4% (footprint 106-3 in Table 1) and +10.3% (footprint 108-4 in Table 2).

Figure 4 show examples of challenging scenarios for automated detection using the

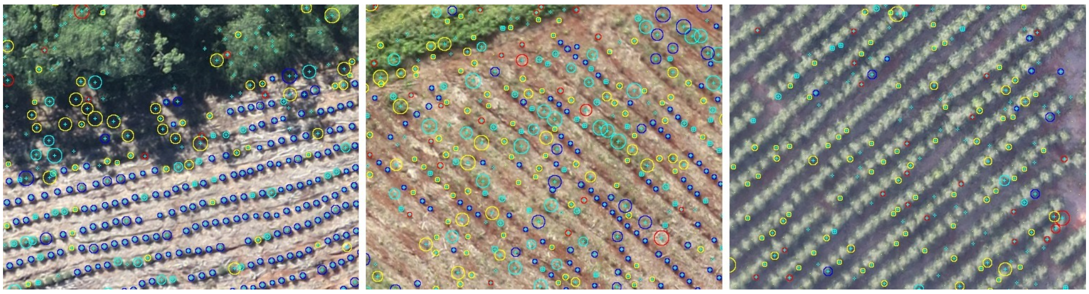


Figura 4: Exemple of challenging scenarios for automated detection. (Left) Shadows cast from an adjacent natural forest. (Middle) Very small trees and possible competition vegetation. (Right) Old forest with overlapping tree crowns.

proposed algorithm. It includes occlusions due to shadows from an adjacent forest, the presence of very small trees and possible nearby competing vegetation, and an old planted forest where most of the adjacent crowns overlap. In this last case, the proposed algorithm is not applicable because the underlying hypothesis that the external rim of the synthetic tree crown template lies on the image background is violated. Further methodological developments are needed to couple with these challenging scenarios.

4. Conclusions

Our approach estimated the location of individual eucalyptus crowns in the early stages of development with accuracies above 90% using RGB imagery with a pixel spacing of about 11 cm. The proposed algorithm seems to be capable of analyzing crowns that are clearly visible in the scenes (preferably larger than 6 pixels wide) when the background of the image is rather homogeneous, as in cases of exposed soil or debris with a discernible brighter color. The automated counts for each tested footprint were within $\pm 11\%$ of the visual estimate, or within $\pm 4\%$ when the footprints were aggregated. Imposing a circular crown shape a priori for multiscale detections seems to be a promising approach and could potentially help to analyze adjacent overlapping crowns. Testing the proposed algorithm on additional data sets and comparing results to census data is needed, as well as further methodological developments to tackle the challenging scenarios identified.

Referências

BREIMAN, L. Random forests. *Machine learning*, v. 45, n. 1, p. 5–32, 2001.

KANG, J. et al. Identifying crown areas in an undulating area planted with eucalyptus using unmanned aerial vehicle near-infrared imagery. *Remote Sensing Letters*, v. 7, p. 561–570, 2016.

KE, Y.; QUACKENBUSH, L. J. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *International Journal of Remote Sensing*, v. 32, n. 17, p. 4725–4747, 2011.

OLIVEIRA, L. T. d. et al. Influence of Eucalyptus sp. stand age on tree counting with LIDAR data. *Cerne*, v. 20, n. 4, p. 557–565, 2014.

POULIOT, D. et al. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, v. 82, p. 322–334, 2002.

SERRA, J.; SOILLE, P. *Mathematical morphology and its applications to image processing*. [S.l.]: Springer Science & Business Media, 2012.

TRIER, Ø. D.; ZORTEA, M.; TONNING, C. Automatic detection of mound structures in airborne laser scanning data. *Journal of Archaeological Science: Reports*, v. 2, p. 69–79, 2015.