Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril 2009, INPE, p. 6797-6804.

A genetic algorithm for citrus tree counting and canopy diameter estimation

João Camargo Neto[1]
José Iguelmar Miranda[1]

Embrapa Information and technology
Caixa Postal 6041, 13083-970 - Campinas - SP, Brazil.
{camargo, miranda}@cnptia.embrapa.br

Keywords: orange tree, citrus counting, satellite images, image processing, genetic algorithm, diameter estimation

Abstract

A new approach for tree citrus identification and counting system from high spatial resolution images, such as Quickbird satellite, was developed. Satellite imagery with low and high density trees was tested to explore the extremes conditions of citrus plantations in southeastern Brazil. The citrus plot boundary was manually delimited and plant's regions were extracted using spectral information. Firstly, last erosion morphological operator was used to initially set up the tree position, after that, the genetic algorithm approach was used to optimize the position and canopy diameter. A complex diffusion filter extracted the tree citrus rows to improve the tree position adjusting over the plantation regions and eliminated false trees located outside tree rows.

Results showed that the genetic algorithm approach optimized the process of identification and counting citrus tree. Also the new approach accommodated the solution for different grove conditions such as plant canopy size, tree rows and planting density.

1. Introduction

The objective of this work is to present the GeoCitrus system, developed in Java, for automatic counting of citrus tree. The work was performed experimentally, in specific areas of interest, through the implementation of several algorithms of digital images processing and a genetic algorithm.

The Brazilian citrus production chain has taken a leading role in the adjustment of its production system to meet the changes of the world economy. For Brazil, the chain citrus sum annually more than $ 1 billion, appearing as one of the main products of exportation or commodities (Neves & Marino, 2002). To export this amount, Brazil has become the world's largest producer of orange, holding 36% of the total produced worldly. However, this production is distributed unevenly between the Brazilian states. For example, in the State of Sao Paulo, the citrus industry is the basis of the economy of 320 municipalities, it is estimated that the orchards have been reduced by approximately 23 million trees. Given that most of the reduction occurred by migration of producers for other more profitable crops (Farias et al., 2003). For the authors, between 2000 and 2001, the sugar cane invaded 80 thousand hectares of former orange groves in the State of Sao Paulo. Fortunately, the Brazilian agricultural sector can count currently with geotechnology tools that can help one to account for his stock of trees. The traditional counting system by sampling and consultation has proven useful and has met the demand of the citrus market. However, this system can be improved significantly.

The improvement of modern space technologies, with the availability of high-spatial resolution satellites, such as Ikonos and QuickBird allowing a new approach to the problem of counting of stock plants and even forecasting of harvest. With such high spatial resolution, it is possible to identify individual trees of orange.

The study of perennial crops from orbital images with spatial resolution of 20 meters or higher offers restrictions. For example, Gordon (Gordon et al., 1986) reported a study with the Landsat TM images in an attempt to discriminate orchards by type of fruit (apple, cherry, pear and peach) and isolate the orchards as a class. The authors found that it was not possible to undertake such a task, because of the difficulty in discerning the reflectance between young and senescent orchards. The only possible differentiation was among orchards and forest vegetation.

For Sanches (Sanches et al., 2005), the study of the citrus orchards with sensors with spatial resolution equivalent to the Landsat TM, requires the understanding of various factors such as changes in the substrate, height and coverage of the land by the trees, geometric features in the provision of plant and the conditions of observation of the sensor. These variations contribute to hinder a standard for the spectral responses of these targets.

Sanches (Sanches et al., 2005) used images from the China-Brazilian satellite, CBERS, CCD sensor, to study the possibility of such images to discriminate varieties of citrus in the state of Sao Paulo. The conclusion of the authors was that the images with spatial resolution of 20m, is useful only to classify citrus grouped into classes. Therefore, discrimination of individual varieties is not possible to be achieved with such images.

The development of software for counting plants, and, consequently, for crop forecast, is strategic for any country. Images of high spatial resolution need to be better studied, mainly, for this purpose. It is necessary to explore new techniques for the processing of digital images which result in a qualitative breakthrough in the area.

According to Martin (Martin, 2003)., the Florida Department of Citrus (FDOC), was funding a pilot project at the National Agricultural Statistics Service (NASS) for the counting of plants using images of high spatial resolution of the QuickBird satellite. Bohac (Bohac, 2005) shows that the previous project progressed and cites an agreement between NASA, the Department of Agriculture of the United States (USDA), the Florida Department of Citrus (FDOC) and other partners who will make a contribution of $ 1.3 million of dollars on a program to use satellite imagery of high resolution for counting of orange trees in Florida and Brazil. In addition, the program provides improvement in the methodology of crop forecasting and learns more about the Brazilian citrus industry. Prices policy of orange's productive chain is based, in part, on the estimated harvest.

Fletcher (2005) evaluated the use of high-resolution aerial photographs to detect citrus orchards, affected by fungi (sooty mould - Capnodium citri), as an indicator of infestation of insects in orchards of citrus. The photos used had almost the same spatial and spectral resolution of the images generated by the QuickBird satellite. The author concluded that satellite imagery of high spatial resolution can be used to detect orchards infected by fungi.

2. Materials and Methods

In this research, a QuickBird satellite images from citrus farm was acquired. A citrus counting process was designed to identify and to estimate canopy diameter of individual orange tree. The algorithm was divided into two main phases, the image processing phase and the optimization phase (Fig.1). In the image processing phase, the plant region was extracted, and based on this information; the citrus rows and the initial tree positions were extracted. In the second phase, the optimization phase, the genetic algorithm was implemented to search for the best configuration of tree position and canopy size. The tree rows and the detected initial tree positions were used as input for the genetic algorithm.

<u>The plant region extraction</u>

The plant region extraction was implemented as a user-driven process wherein the color image was segmented into a set of spectral classes using a clustering K-means algorithm (Spaeth, 1980). Using a set of spectral samples of the citrus tree, (Fig. 2), the classes of plant region were separated from bare soil, tree shadow, weeds and agriculture residue, (Fig. 3). In these experiments, we used 6 classes for the K-means algorithm.

<u>The extraction of tree rows</u>

The extraction of tree rows was accomplished using the complex diffusion filter (Gilboa et al., 2004), that generates a bluer (???) connected region for the tree rows and dark regions between tree rows. With Zadeh's fuzzy intensification technique (Ross, 1995) the tree rows regions were enhanced. The morphological skeleton algorithm was used to generate the tree rows. These processes are showed in Figure 3.

<u>Initial tree position detection</u>

The initial tree canopy position was extracted using the morphological operator last erosion (Dougherty, 1992), that finds for each connected component the last subsets removed by the morphological erosion (it corresponds to the local maximum of the distance map). This operation not only generated a unique point for each tree but also separated connected trees. Using the tree rows information, the false trees between the tree rows could be eliminated (fig 4).

Genetic Algorithm design

Unhappily, the position of some trees was not always generated at the center of canopy; some tree centers were missed; and some tree centers were generated too close to another tree center. To solve this kind of problem, a genetic algorithm, introduced by Holland (Holland, 1975), was used in the third step to optimize search for the optimum tree position and estimate of diameter, (Fig. 5).

<u>Chromosome</u>

The chromosome represents the solution of the problem, the citrus groves with a set of trees defined as a triplet, the position (latitude and longitude) and canopy diameter information assigned to it (x_1, y_1, d_1). An example of chromosome structure is shown in Figure 6

<u>Population Size</u>

The population is a set of chromosome. For the citrus tree counting application, the population size was set up with a set of 100 chromosomes. The subsequent population (generations) of chromosomes was generated using the genetic algorithm operators such as elitism, crossover and mutation. The elitism was set to 4 chromosomes. This operator kept the chromosomes with the best fitness value in the next generation by copying them to the new population.

<u>Selection</u>

In order to generate offspring's chromosome, two parents chromosome were extracted from the population, based on its fitness value using the roulette wheel method, for recombination in crossover operation.

<u>Crossover and Mutation</u>

Crossover and mutation operator generate the next population from the genetic material of parents. The crossover generated two offspring using a random single point selection method where the two parent chromosomes were selected by the roulette wheel method. The selected chromosomes were divided and the piece of each chromosome was recombined to generate the offspring's chromosome. The crossover probability was set to 85%. The mutation operator was used to change the chromosomes resulting from the crossover due to a change in canopy size, position, new addition, merger or removal of trees. A rate of 10% was set up for mutation operator.

Function evaluation

The best solution is reached when the trees cover the entire tree regions without covering any soil region. The fitness value, equation 1, measures how good the chromosomes represent this best solution. This fitness value is the mean between the ***CovFt***, equation 2, which measures the ratio between the regions covered for the all trees in the chromosome, and the total tree rows regions extracted and the ***treesFt***, equation 3, which measure the ratio between the plant region and the soil region covered for all the trees in the chromosome.

$$fitness\_Value = \frac{CovFt + treesFt}{2} \qquad 1)$$

where: $\operatorname{cov} Ft = \frac{1}{CovTotal} \sum_{i=1}^{n} CovTree_i$, 2)

$$treesFt = \frac{\sum_{i=1}^{n} CovTree}{\sum_{i=1}^{n} CovSoil} \qquad 3)$$

Stopping Criteria

The genetic algorithm stopped when one of the following three conditions was satisfied. First, the number of generations was greater than 100. Second, it reached the threshold for acceptable citrus tree area, covering 95% of the plant region. Third, there was no fitness value improvement after 20 generation.

The accuracy of the method for different grove conditions was obtained using the following equation:

$$acuracy = \frac{GeoCitrus\_Count}{Visual\_Count + double\_trees + mis\sin g\_trees + false\_trees} \qquad 4)$$

Where *GeoCitrus_Count* is the result for the computerized tree count, *Visual_Count* is a manual tree count obtained from the image, *doube_trees* is the double count for the same tree, *missing_trees* is the number of trees that the system was not able to count, and *false_trees* is the number of trees that the algorithm generate in position where there is not plant region.

3. Results and Discussion

The GeoCitrus algorithm was tested in different grove conditions as shown from figure 7a to figure 7d. The results are presented in table 1 and 2. Tree types of error were detected; the missing trees, double trees and false trees. The missing trees error occurred where the plant region extraction failed due to canopy shadow or non homogeneous spectral canopy response. In general, double trees error occurred at grove condition with large canopy size. This error was generated initially by the last erosion operator generate a tree position outside of the center of plant region, generating a free plant region. In order to cover all plant regions the algorithm generates a new tree to fill up the free plant space. The double trees error also could happen when the tree boundary is not very well extracted. The last erosion in this case could generate for the same plant region two points that will represent for the algorithm two small trees. False trees occur when some weed region between trees is misclassified as plant region during the plant extraction region phase. Also there are some tree positions and size errors due to genetic algorithm approach. The genetic algorithm optimizes the set of the trees, that is, the overall result and not tree by tree. The result represents the best fitness value for set of trees and it can have some small errors.

The accuracy was measured for it grove conditions and is presented in table 2. For the uniform tree size condition the algorithm reached a mean of 97% accuracy and for different size and space between the tree rows the algorithm reached a mean of 93%. The algorithm shows some difficult to find the best configuration when tree size was not uniform. Also the table 2 presents the percent of canopy cover. The mean 82% was reached, but more experiments have to be done to compare algorithm results with the real trees canopy area on the field.

4. Conclusions

The GeoCitrus system showed the ability to count trees and measure the canopy diameter in different grove conditions. Using genetic algorithm optimization approach, the system was able to adjust the canopy center position and also estimate the canopy diameter. There are many improvements that can be done, but even that, the system can be used to count and monitor the citrus grove.

5. Literature review

Bohac, N. High resolution imagery to aid citrus industry. Earth Observation Magazine, vol. XIV, no. 2, April, 2005. Site: <http://www.eomonline.com/EOM_Apr05/departments02.html>.

Dougherty, E. An introduction to morphological image processing. SPIE Optical Engineering Press. 1992.

Farias, P. R. S.; Nociti, L. A. S.; Barbosa, J. C.; Perecin, D. Agricultura de precisão: mapeamento da produtividade em pomares cítricos usando geoestatística. Revista Brasileira de Fruticultura, v. 25, n. 2, p. 235-241, 2003

Fletcher, R. S. Evaluating high spatial resolution imagery for detecting citrus orchards affected by sooty mould. International Journal of Remote Sensing, v. 26, n. 3, p. 495-502, 2005.

Gilboa, G.; Soche, N.; Zeevi, Y. Y. Image enhancement and denoising by complex diffusion processes. IEEE Transactions on Pattern Analysis and Machine Intelligence. 25(8):1020-1036, 2004.

Gordon, D. K.; Philipson, W. R.; Philpot, W. D. Fruit tree inventory with Landsat thematic mapper data. Photogrammetric Engineering and Remote Sensing, v. 52, n. 12, p. 1871-1876, 1986

Holland, J.H. Adaptation in natural and artificial system. Ann Arbor, The University of Michigan, Press, 1975.

Martin, J. Eyeing the competition. Citrus & Vegetable Magazine. Site: < http://www.citrusandvegetable.com/home/2003_AprilSatellite.html

Neves, M. F.; Marino, M. K. Estudo de competitividade de cadeias integradas no Brasil: impactos das zonas de livre comércio. Unicamp, 2002.

Ross, T.J.,. Fuzzy Logic with Engineering Applications. McGraw-Hill, New York, 1995

Sanches, I.D.; Gurtler, S.; Formaggio, A.R. Discriminação de variedades de citros em imagens CCD CBERS-2. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 12., Goiânia. Anais... São José dos Campos: INPE. p. 277-284. CD-ROM. ISBN 85-17-00018-8, 2005

Spaeth , H. Cluster Analysis Algorithms for Data Reduction and Classification of Objects. Ellis Horwood, 1980.

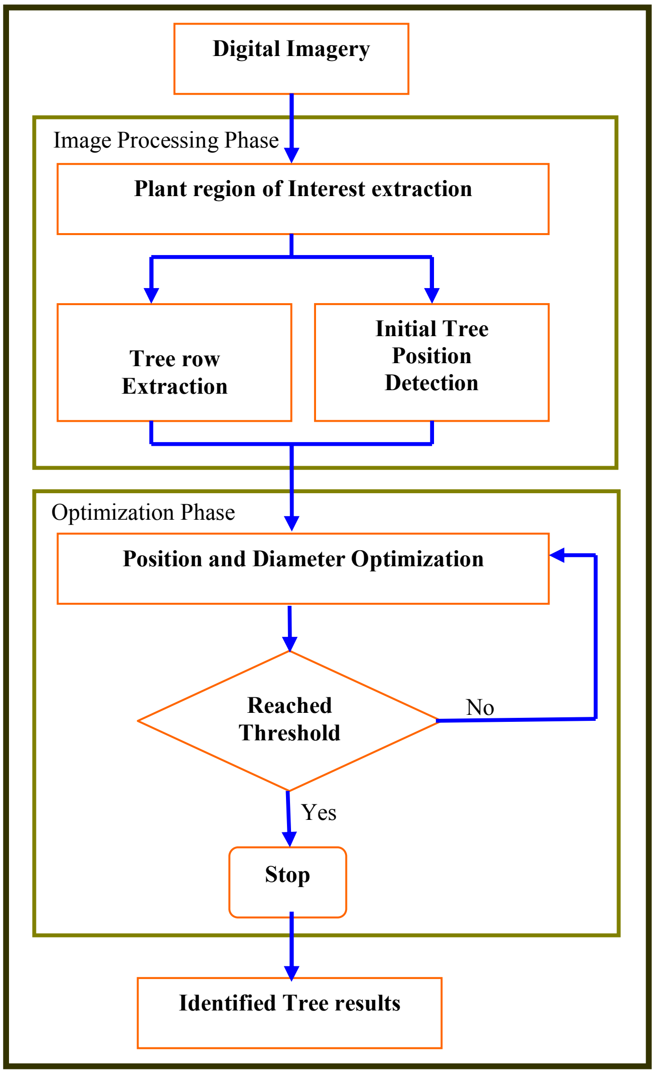
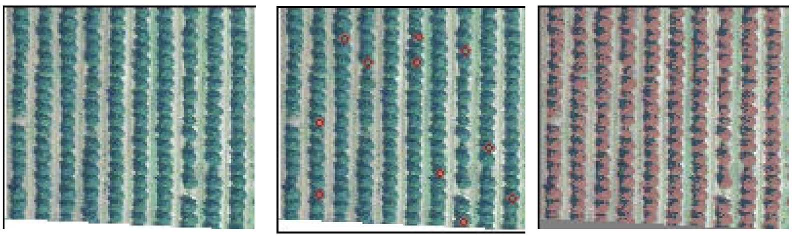
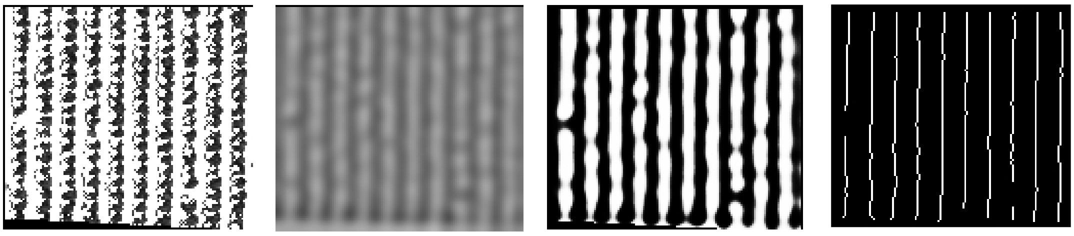


Figure 1. Flow chart of digital image analysis procedure for tree count and diameter estimation.



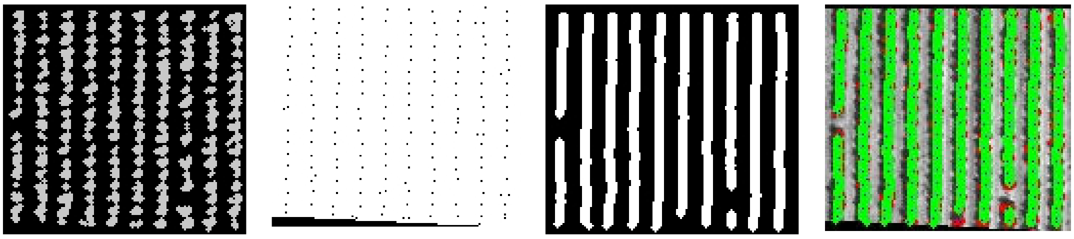
a) b) c)

Figure 2. Plant region of Interest extraction; a) original image, b) tree samples extraction, c) plant region extracted



a) b) c) d)

Figure 3. Tree rows extraction; a) plant region, b) complex diffusion filter results, c) Zadeh intensification technique, d) Morphological Skeleton operator



a) b) c) d)

Figure 4. Initial tree position detection; a) Morphological Open operator, b) last erosion operator, c) tree rows dilation operator, d) elimination of false trees

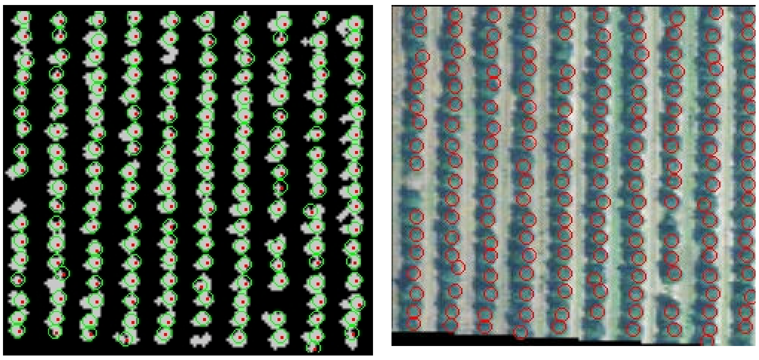


Figure 5. Initial tree configuration

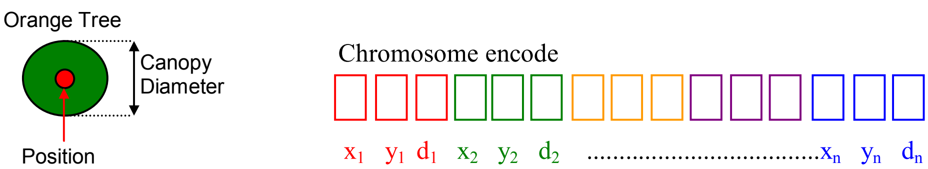


Figure 6. Chromosome encoded as a set of triplex, the position (x, y) and canopy diameter.

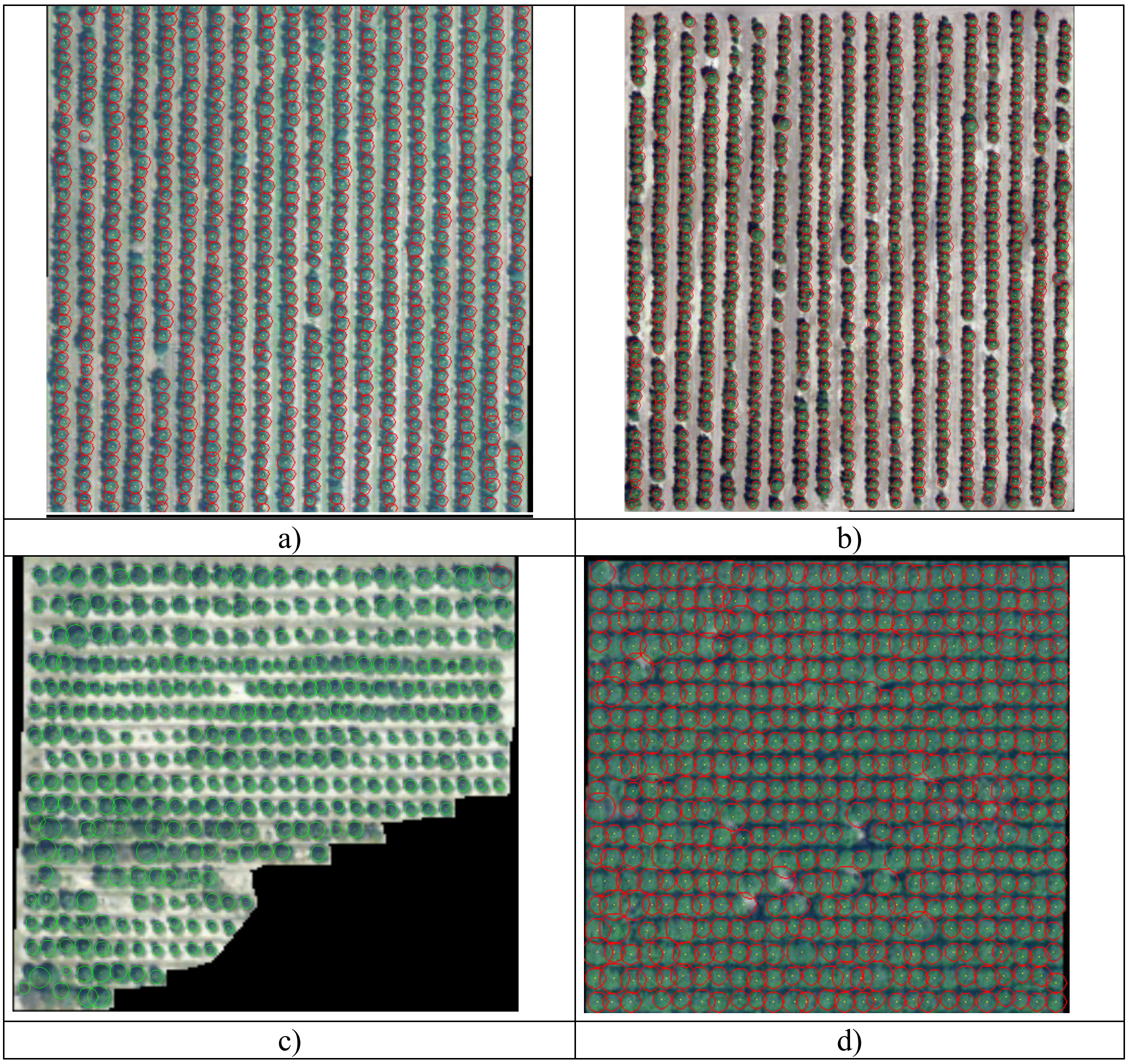


Figure 7. image samples: a and b) uniform tree size and spaced tree rows, c) different tree size with low tree row space, and d) different tree size with no space between tree rows.

Table 1. Results of visual and GeoCitrus tree count

Figure	Grove Condition		Tree Count		Error		
	Tree rows space	Tree size	Visual	GeoCitrus	Missing trees	Double trees	False trees
7a	high	uniform	754	747	12	0	0
7b	high	uniform	893	885	13	1	4
7c	low	different	356	343	14	0	1
7d	without	different	484	482	13	8	3

Table 2. GeoCitrus cover area estimation and accuracy

Figure	Percent of canopy cover	Accuracy
7a	0.78	0.97
7b	0.80	0.97
7c	0.82	0.92
7d	0.87	0.95