ALMOND CANOPY DETECTION AND SEGMENTATION USING REMOTE SENSING DATA DRONES

Tiebiao Zhao[1], Merari Cisneros[1], Qi Yang[1, 2], Yanzhu Zhang[1, 2], and YangQuan Chen[1]

[1]Mechatronics, Embedded Systems and Automation (MESA) Lab, School of Engineering, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA

[2]School of Mechanical Engineering, ShenYang LiGong University, Shenyang 110159, China

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract.

The development of Unmanned Aerial System (UAV) makes it possible to take high resolution images of trees easily. These images could help better manage the orchard. However, more research is necessary to extract useful information from these images. For example, irrigation schedule and yield prediction both rely on accurate measurement of canopy size. In this paper, a workflow is proposed to count trees and measure the canopy size of each individual tree. The performances of three different methods to classify tree canopies are compared. Then morphological methods are used to filter grass patches and separate the trees from each other. Finally, the number of trees and the size of tree canopies are obtained.

Keywords. *Tree counting, canopy size measurement, unmanned aerial system (UAV)*

1 Introduction

There are quite a few studies on tree inventory. Jang (2008) discussed how to retrieve raster height of apple tree groves using LIDAR. Karantzalos and Argialas (2004) obtained robust counting results for well-spaced, non-overlapping olive trees based on aerial cameras. Camargo and Miranda (2009) researched a method to count citrus trees and measure the diameters based on satellite images. She et al. (2014) proposed a method to count nursery crops based on aerial images. None of these studies provide an approach to count almond trees and estimate the size of tree canopies with the help of unmanned aerial system (UAV).

It has been shown that the size of canopy is related to crop coefficient (Johnson and Scholasch, 2005) and can be used to optimize irrigation schedule. Motivated by this, we are interested to count trees and more importantly, to measure the size of tree canopy of each individual tree.

2 Materials and Methods

2.1 Field Description

This test field is a mature, commercial almond (Prunus dulcis) orchard in Ballico, Merced County, California (37.493498°N, -120.634914°W). Three different varieties were planted on Lovell Peach rootstock 15 years ago, spacing at 5.5m*6.1m.

2.2 UAV and Payload

The UAV used in this study was built using DIY Quadkit (3DRobotics, Berkeley, USA), modified to carry a single camera payload to do the remote sensing. The maximum takeoff weight of the aircraft is 2.0 Kg and its flight time is 15 minutes. One modified commercial-off-the-shelve (COTS) near-infrared (NIR) camera (ELPH110HS, Canon, Japan) was flown from August, 2014 to October, 2015. It includes three bands, NIR, green, blue. It has a resolution of 4608*3459, with 24 bit radiometric resolution and focal length of 4.3 mm. The aerial image was stitched using PhotoScan (Agisoft, Russia).

2.3 Methods Description

2.3.1 Ostu Method

In 1979 Noboyuki Otsu devised a way of easily evaluating the effectiveness of a threshold to produce an automated means of threshold selection (Otsu, 1975). This variant of clustering creates tight clusters in the hopes of preventing overlap. This is done by altering thresholds, but by increasing the spread of one threshold results in the decrease in another threshold. As a rule, one wants to choose the threshold that will minimize the combined spread. Through computational analysis Otsu was able to come up with the between class variance which could be simplified to:

$$\sigma^2{}_{Between(T)} = n_B(T) n_o(T) [\mu_B(T) - \mu_o(T)]^2 \quad (1)$$

Since the computation for the candidacy of thresholds are not independent of each other when moving from threshold to threshold the $n_B(T), and\ n_o(T)$ values can be updated along with their cluster means $\mu_B(T), and\ \mu_o(T)$ as pixels change location from one cluster to another. Through the use of recurrence relations, the aforementioned values can be updated in order to test another threshold. In short, Otsu's method provides a powerful tool for threshold selection which facilitates image segmentation.

2.3.2 HSV color space

Hue, Saturation, and Value (HSV) is a three dimensional color space which sees use primarily in the generation of computer graphics. In the HSV space Hue represents the color in an image represented in some form such as an angle, typically in the 0^0 to 360^0 range. Saturation denotes the

span of grey in the HSV color space, usually represented by a range of 0% to 100% where 0% represents the color grey, 100% a primary color, and anything in between is a mixture of the two. Value is the brightness of a color which is dependent on the amount of saturation present. It too is usually within the scale of 0% to 100% in which 0% corresponds to a hue that is completely black while one with 100% gives very bright hues. By converting images from the RGB plane into the HSV an object's specific hues, saturation, and Value or intensity in a given image can be exploited to isolate the selected image via color segmentation and a threshold mask based either off Hue, Saturation, or intensity (Gonzalez and Woods, 2008).

2.3.3 Gray Level Concurrence Matrix (GLCM)

The unique texture of an image in question can be harnessed to distinguish it from the textures of other objects or even the background when combined with the K-Nearest Neighbor algorithm (Guru et al., 2010). This twostep process first requires the use of the Gray Level Concurrence Matrix method. This method extracts textural features from a given image and stores them in matrix that has the same number of rows and columns that a given image has gray levels (Albregtsen, 2008). The GLCM shows how often two pixels i and j representing intensity i and j respectively occur at a given pixel distance ($\Delta x, \Delta y$) within a neighborhood denoted by the matrix element $P(i, j|\Delta x, \Delta y)$. Additionally, the matrix element $P(i,j|d,\theta)$ accounts for changes in the gray levels in i and j at a distance d positioned at angle θ. Haralick et al. (1973) developed set of features for classifying pictorial data. Specifically, it provides a general technique for the extraction of textural properties from gray-tone special-dependency matrices. Haralick, et al. created 14 measures for the extraction of textural features. However, Newsam and Kamath (2004) showed that only about five are frequently used such as Angular Second Moment (ASM), Contrast (COR), Inverse Different Moment (IDM), Entropy (ENT), and Correlation (COR) defined by their respective equations.

$$ASM = \sum_{i,j} P(i,j) \tag{2}$$

$$CON = \sum_{i,j} |i-j|^2 P(i,j) \tag{3}$$

$$IDM = \sum_{i,j} \frac{P(i,j)}{1+|i-j|^2} \tag{4}$$

$$ENT = \sum_{i,j=0}^{N-1} -\ln(P_{i,j}) P_{i,j} \tag{5}$$

$$COR = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\vec{P}(i,j)}{\sigma_i \sigma_j} \tag{6}$$

2.3.4 K-NN

In a paper by Fix and Hodges (1952) an important method regrading nonparametric discrimination through the use of consistency properties was conceived. This method deals with, at least in a two population case, a random variable Z with value z that is present within some space that is p dimensional based of the distribution of the two populations denoted by F and G. It is possible to discriminate between the two populations with training samples (X1,X2,…,Xm) and (Y1,Y2,…,Ym) from populations F and G respectively. The density in each population is denoted f and g. If the values of population F and G are known, then the classification of a sample point at z is given by the likelihood ratio:

$$\iota(z) = \frac{f(z)}{g(z)} \tag{7}$$

However, if the values of the given populations are unknown then it is commonly assumed that f and g are part of a parametric family in which estimates of f and g can be made for insertion into equation

(7). Fix and Hodge's method instead assumes that f and g satisfy specific smoothness assumptions (Silverman et al., 1989). From this observation Fix and Hodge were able to devise two methods of nonparametric discrimination through the use of density estimates one of which is the variant of nearest neighbor density estimation used in this study. This nearest neighbor variant method makes use of the naïve kernel density estimate given by:

$$\hat{f}(z) = m^{-1} \sum_{i=1}^{m} K_m(X_i - z) \quad (8)$$

to produce an estimate for density f by choosing a neighborhood ∆m that is just large enough to encompass a given number of k points in a sample. The new neighborhood now denoted by ∆m,n aids in the creation of an estimate for density g. Eventually, these values are used to create a nearest neighbor regression estimate based of the value of k, where k denotes the number of nearest neighbors to include in the discrimination process. The varying values of k determine what class a specific point in question will be assigned. A k value of 1 will classify this point based on its nearest neighbor while a k value that is larger than 1 will base the classification off which neighbor is present in majority of all possible neighbors. In this manner the K-Means Nearest neighbor provides a powerful method for object classification.

3 Results and discussion

The test image for this paper was taken in May 20th, 2015, as show in Fig 2.(a). In this paper we tried two types of classification methods to differentiate trees from non-tree parts. One of the methods is unsupervised the Otsu threshold method. The other two are supervised methods based on histogram similarity in HSV color space and text features described by GLCM. 20 samples are prepared as model samples for supervised classifications, including 10 tree samples, 2 shade samples, 3 soil samples and 5 grass samples. All these samples are of size 10*10. It is clear that grass and trees have almost the same color, so the grass samples account for 50 percent of non-tree samples. All the 20 samples are displayed in Fig. 1. In addition to 20 model samples, there are another 100 samples selected to test the classification performance, composed of 50 tree samples and 50 non-tree samples.

3.1 Classification

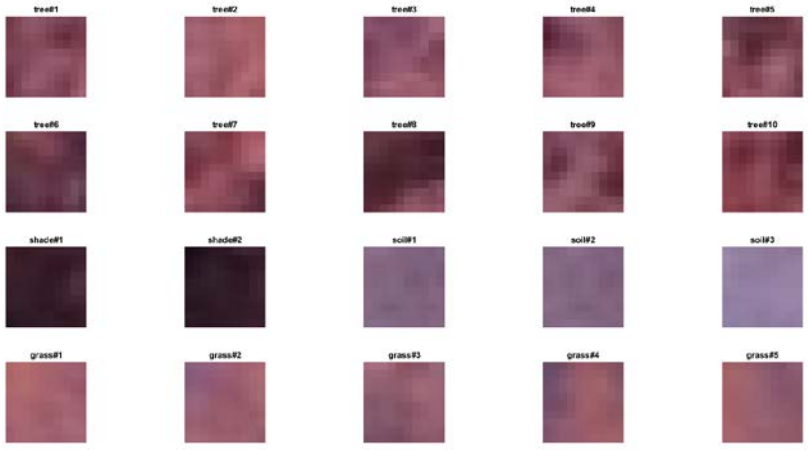

Fig 1. Model Samples, including 10 tree samples, 2 shade samples, 3 soil samples and 5 grass samples.

3.1.1 Otsu method

According to the fact that the difference of crops between NIR and blue is larger than that of soil, the original image is converted to a gray image by taking the difference between NIR band and blue

band. The threshold is calculated based on the large image Fig.2(a) using the Otsu method. For each test sample, if the number of pixels, whose intensity are larger than the threshold, is greater than 50, then this sample is classified as a tree sample. Otherwise, it is marked as a non-tree sample. According to the results (Table 1), all the tree samples are predicted as trees. The problem is that, however, more than 40 non-tree samples are recognized as trees, which will introduce a big trouble to measure the size of tree canopies.

Table 1. Classification performance of Ostu method.

Ostu (grey image)	Tree(Predicted)	Non-tree(Predicted)
Tree(Real)	50	0
Non-tree(Real)	40	10

3.1.2 HSV histogram

The RGB color space is nonlinear and it is better to convert the image from RGB to more uniform color space HSV. First all the samples are converted to HSV color space. Then, histograms of all the samples in H, S, V are calculated individually. The histograms are divided into 32 bins and each bin cuts 3.125 (1/32) percent of the range of H, S, V intensity. The empty bins are set to one to make the definition of distance between histogram distributions applicable according to Kullback's minimum cross-entropy principle (Equation. (9)). The distances of H, S, V are added together as the overall distance between model samples and test samples. Finally, K-NN is applied to predict the label of the sample, where k is 5 in the test.

$$D(s,m) = \sum_{i=1}^{n} s_i log \frac{s_i}{m_i} \quad (9)$$

It shows that (Table 2) HSV method helps filter non-tree pixels, where only 10 non-tree samples are labeled as trees, although it increases the possibility to classify tree pixels as non-tree pixels at the same time. Further check indicates that all the non-tree samples recognized as tree samples are grass samples, which means grass and trees have similar color even in HSV space.

Table 2. Classification performance of HSV method.

HSV	Tree(Predicted)	Non-tree(Predicted)
Tree(Real)	45	5
Non-tree(Real)	10	40

3.1.3 GLCM+HSV histogram

In order to get better classification accuracy, we also combined texture with color to help differentiate grass from trees. As an important texture feature, GLCM has been widely used in many applications. Gray-level co-occurrence matrixes are calculated in 8 directions, $0°, 45°, 90°, 135°, 180°, 225°, 315°$, at 16 gray levels in NIR, red and blue bands. Contrast, correlation, energy and homogeneity obtained based from the matrixes and then are averaged in 8 directions to become rotation invariant. For each sample, there are three GLCM feature vectors composed of contrast, correlation, energy and homogeneity in NIR, green and blue bands. The GLCM distance between samples is measured by sum of square of distance between their GLCM vectors in three bands. The HSV distance and GLCM distance are normalized individually before added as a combined feature. Similarly, K-NN is applied in the tests. As shown in Table 3, the combined method decreases the fault rate of classifying non-tree as trees, which is better than both HSV method and Otsu method. At the same time, all the tree samples are recognized, better than HSV method, and as good as Otsu method.

Table 3. Classification performance of HSV combined with GLCM.

HSV+GLCM	Tree(Predicted)	Non-tree(Predicted)
Tree(Real)	50	0
Non-tree(Real)	7	43

3.2 Segmentation

The original image is converted to binary image using three methods, the Otsu method, the HSV method and the combined method. HSV method does help filter out some grass patches, as pointed out by red arrows in Fig.2(b,c). However, the combined method recognizes many soil parts as canopies. Its performance is not as good as the one in the previous test. Considering the requirement of learning samples, and border effect of texture methods, we finally chose the Otsu method. The Otsu method is unsupervised, so it does not need training samples. As for the high fault rate of classifying grass as trees, we can tell the difference based on their morphological features.

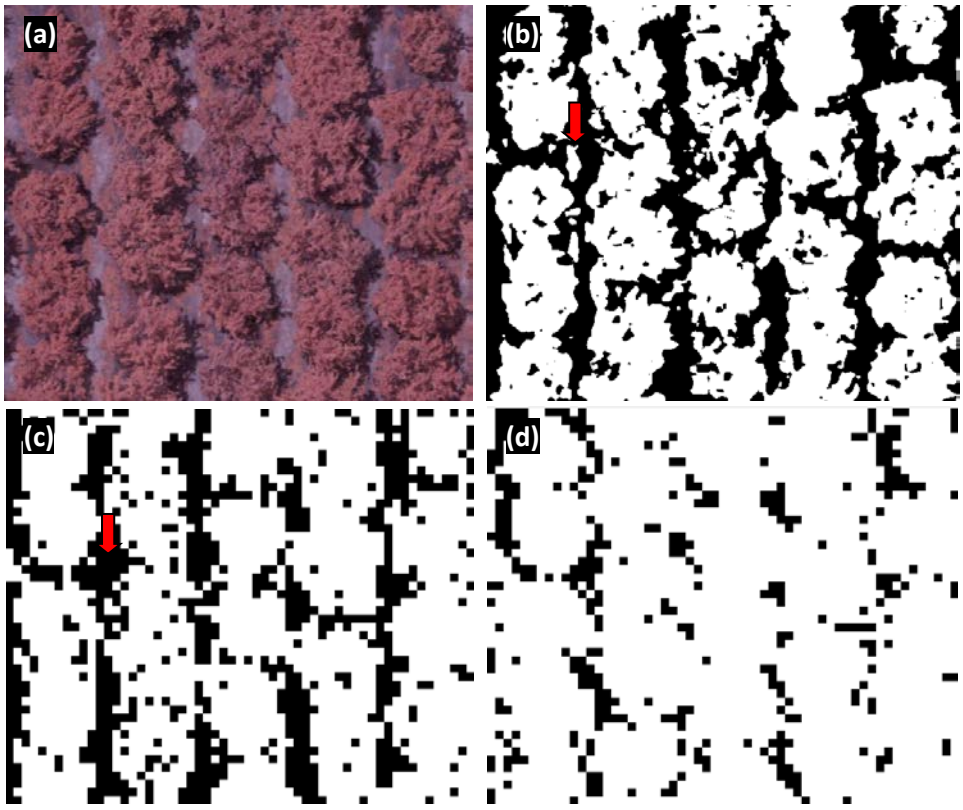

Fig 2. (a) Original image. (b) Binary image with Ostu method. (c) Binary image with HSV method. (d) Binary image with HSV combined with GLCM method. Tree canopy is white and others are black.

3.2.1 Filter small grass region and correct shaded region

There are some isolated patches labeled as canopies in the image obtained directly from the Otsu method. These patches are either small grass regions or soil regions, which can be deleted based on their size. Fig.3(b) shows the image obtained after white patches with area less than 500 pixels are deleted. In addition, there are some patches, shaded region in the canopies are labeled as non-tree. These patches can be corrected according to their size too. Fig.3(c) shows the image after the dark patches with area less than 500 pixels are re-labeled as tree canopies. Note that erosion and dilation are not appropriate here, since they will change the size of canopies. In this paper, we are interested in measuring the size of canopies as accurate as possible.

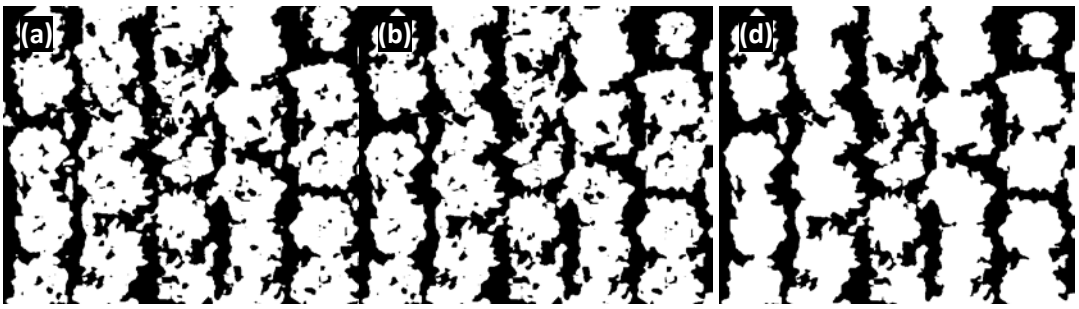

Fig 3. (a) Binary image obtained with Ostu method. (b) Binary image after deleting small patches outside canopies. (c) Binary image after deleting small patches within canopies.

3.2.2 Delete large grass region

Small isolated patches are easy to process according to their size. However, there are still some large grass patches left connected with tree canopies. We have to cut these patches away from canopies according to certain features. First, they are narrower than canopies. Fig. 4(a) filters all the patches less than 25 pixels wide. The problem is that, two type of canopy region are filtered too. One part is from the top or bottom of the canopy circle. The other part is from the region connecting two canopies. These two parts are characterized by their orientations, height and major axis length. Fig. 4(b) shows the image with patches height less than 5 pixels. After patches of orientation angles (absolute value) less than 5 degrees, and major axis length less than 25 pixels are taken away, the obtained image is as Fig. 4(c), where all the large grass patches are left. By taking all these large grass patches, we obtain the image in Fig. 4(d). There are two remaining troubles preventing counting trees and the pixels in each canopy. First, there are certain parts between the forth row and the fifth row are connected. Second, there are many tree canopies are connected in each row.

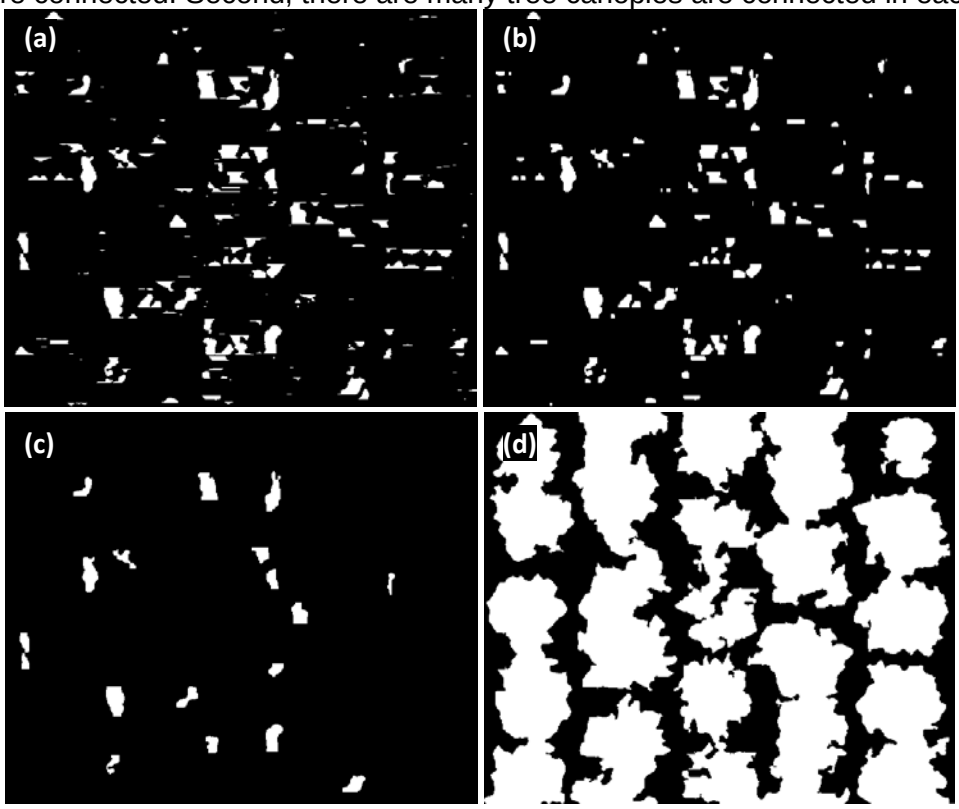

Fig 4. (a) Connected regions less than 25 pixels in the horizontal direction. (b) Image 4.(a) after deleting regions less than 5 in vertical direction. (c) Image 4.(b) after deleting regions with orientation angle (absolute value) less than 5 degree, major axis length less than 25 pixels.

3.2.3 Separate every row

To separate the connected rows, we need to find the border for each row. By calculating the number of pixels in each column, we can see clear six local minimum points in Fig 5(a). In Fig 5.(b), a line is put between rows right at the point with locally minimum number of canopy pixels. It turns out these lines are good estimations of borders to separate each row. After deleting the canopy pixels in the lines and the remained isolated patches generated, we have five separate rows in Fig 5.(d).

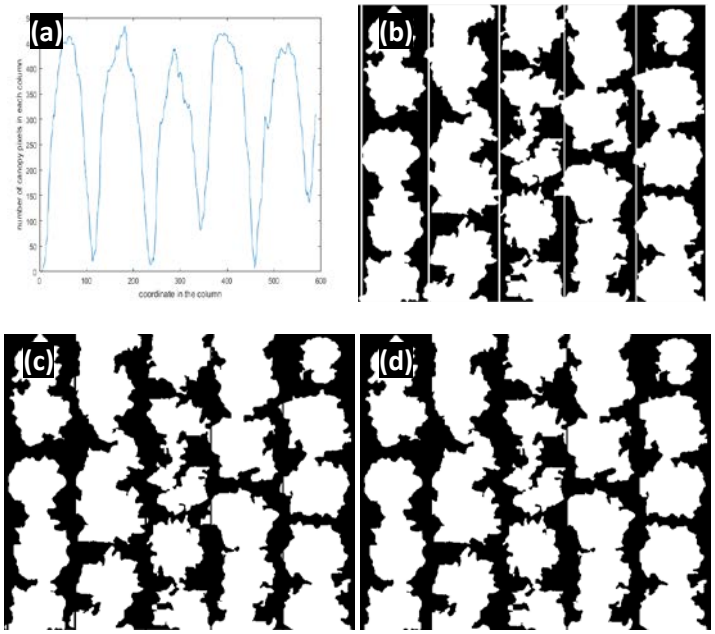

Fig 5. (a) Number of canopy pixels in each column. (b) Binary image with lines separating each row of trees. (c) Pixels on the lines marked as non-canopy. (d) Image 5.(c) after deleting regions smaller than 100.

3.2.2 Separate every tree

After obtain the border for each row, we could focus on separating each tree in a row. Similarly, we can segment each tree in a row according to the number of canopy pixels in each line within a row. Fig 6 shows the number of pixels in each line in each row, where there is a local minimum every around 100 lines. After deleting the pixels in the lines where these local minimum are, we could have the image Fig 7.(a) composed of 27 single tree canopies. In both Fig 7.(a) and Fig 7.(b), centroid of each tree canopy are marked in blue. The location and area (in pixels) for each individual tree are listed in Table 4.

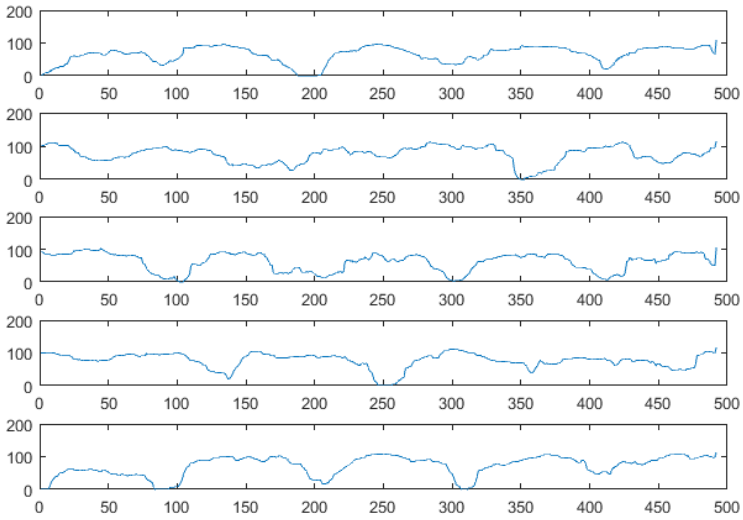

Fig 6. Number of canopy pixels in each line within a row of trees.

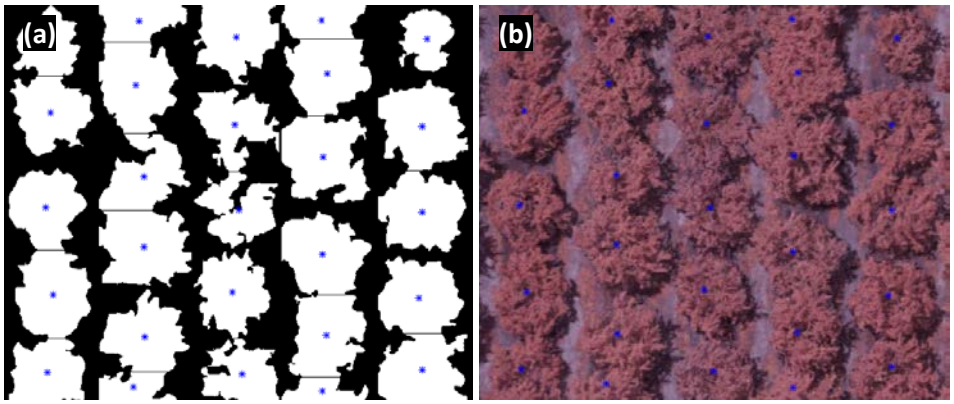

Fig 7. (a) Each tree is separated and its centroid is marked in the binary image. (b) The centroid of each tree is marked in the original image.

Table 4. Centroid location and the area of each tree.

Tree NO.	Centroid		Area
	Row	Column	
1	251.1815	53.08843	6638
2	455.0114	55.39915	5875
3	134.3472	58.93714	6602
4	50.42477	55.93564	4739
5	358.6278	61.64048	7677
6	20.59761	167.9282	3929
7	99.66654	164.2116	8166
8	212.7864	174.4781	6331
9	300.0748	173.6795	8424
10	410.5728	175.4413	6850
11	473.347	160.7197	2519

12	40.59725	287.8382	7044
13	148.8681	285.6326	5776
14	254.0608	290.8541	5311
15	356.4398	283.2237	6473
16	457.3133	294.7792	5036
17	20.20303	392.8246	3694
18	189.4533	395.4826	8649
19	309.284	394.2629	7899
20	477.9896	393.8793	2021
21	85.88611	399.8406	7551
22	409.167	399.3958	7448
23	150.6768	516.7222	8484
24	256.7604	516.6707	7910
25	362.5341	513.098	7483
26	452.3773	516.5105	7170
27	43.06107	522.5649	3668

Conclusion or Summary

Three different algorithms are developed in the paper to classify trees from grass and soil. Tests show that the combined method of HSV and GLCM performs best, but when it is applied to the original image, more non-tree pixels are misclassified as trees. Therefore, the Otsu method is used to obtain the binary image for post processing. Then a post-processing workflow is developed to segment the connected trees, count the number of trees, and measure the size of the canopies. Though the hypothesis for this approach that trees are uniformly spaced is general, the approach still needs to be tested on more images. For example, more images taken at different growing stages and different weather conditions could bring new troubles, which require further detailed research.

Acknowledgements

This work is supported in part by UC ANR Competitive Grant Award No. 13-2628(2014-2019) entitled "Evaluating and extending the use of small, multi-rotor unmanned aerial vehicles (UAV's) as a crop monitoring tool."

The authors would like to thank David Doll, Andrew Ray, Larry Burrow for lending their expertise in almond orchard. Thanks go to the MESA Lab Scientific Data Drone crew members Ph.D Student Brandon Stark, undergraduate researchers Yoni Schemelinin, Andreas Anderson for contributions in conducting flight missions, undergraduate researchers Adrian Hernandez, Roberto Nava for image pre-processing.

References

Albregtsen, F. (2008). Statistical texture measures computed from gray level coocurrence matrices. Image processing laboratory, department of informatics, University of Oslo, 1-14.

Camargo Neto, J., & Miranda, J. I. (2008, January). Orange Tree Counting and Canopy Diameter Estimation with Genetic Algorithm. In International Symposium on Application of Precision Agriculture for Fruits and Vegetables 824 (pp. 29-36).

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(5), 603-619.

Fix, E., & Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. California Univ

Berkeley.

Gonzalez, R. C. & Woods, R. E., (2008). Digital Image Processing M.McDonald, (Ed.) (3rd ed.). Upper Saddle River, NJ: Pearson Education Inc..

Guru, D. S., Sharath, Y. H., & Manjunath, S. (2010). Texture features and KNN in classification of flower images. IJCA, Special Issue on RTIPPR (1), 21-29.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on, (6), 610-621.

Jang, J. D., Payan, V., Viau, A. A., & Devost, A. (2008). The use of airborne lidar for orchard tree inventory. International Journal of Remote Sensing,29(6), 1767-1780.

Johnson, L., & Scholasch, T. (2005). Remote sensing of shaded area in vineyards. HortTechnology, 15(4), 859-863.

Karantzalos, K. G., & Argialas, D. P. (2004, July). Towards automatic olive tree extraction from satellite imagery. In Geo-Imagery Bridging Continents. XXth ISPRS Congress (pp. 12-23).

Morse, B. S. (2000). Lecture 4: Thresholding. Brigham Young University.

Newsam, S. D., & Kamath, C. (2004, April). Retrieval using texture features in high-resolution multispectral satellite imagery. In Defense and Security (pp. 21-32). International Society for Optics and Photonics.

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1), 51-59.

Otsu, N. (1975). A threshold selection method from gray-level histograms. Automatica, 11(285-296), 23-27.

She, Y., R. Ehsani, J. Robbins, J. N. Leiva* and J. Owen. (2014). Applications of small UAV systems for tree and nursery inventory management. Proc. 12th Intl. Conf. Precis. Agr. (ICPA)

Silverman, B., & Jones, M. (1989). E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951). International Statistical Review / Revue Internationale De Statistique, 57(3), 233-238. doi:1. Retrieved from http://www.jstor.org/stable/1403796 doi:1

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In ICML (Vol. 1, pp. 577-584).