State Horticultural Association of Pennsylvania **Research Grant Proposal**

Title: How many flowers are there, really? Autonomous estimation of flower counts with image pairs

Personnel:

Principal Investigator: Dr. Henry Medeiros, Assistant Professor, Marquette University, 1551 W Wisconsin Ave., Milwaukee, WI 53233. (414) 288-6186 henry.medeiros@marquette.edu

Co-principal investigator: Dr. Amy Tabb, Research Agricultural Engineer, USDA Agricultural Research Service (ARS) Appalachian Fruit Research Station (AFRS), 2217 Wiltshire Rd. Kearneysville, WV 25430. (304) 725-3451 x386 amy.tabb@ars.usda.gov

Duration of Project: One year. 9/1/2019-6/1/2020

Justification:

Need for bloom estimates in fruit production. Growers and extension specialists use bloom counts to schedule thinning sprays as a component of crop load management. In apple, the pollen tube model (Yoder et al. 2012) requires bloom counts as an input and so does the protocol for precision chemical thinning of Robinson et al. (2014). For peach, bloom estimates have applications for precision crop load management, as well as automated thinning, such as with string thinners (Auxt Baugher 2010) or other selective means such as robotics (Lyons et al. 2015). The current practice in assessing bloom is human sampling.

Drawbacks of current practice. Human-performed sampling has many drawbacks, such as: 1) Tree inspection is time-consuming and hence the number of trees that can be inspected by the grower is limited; 2) Estimation of the blooming intensity or fruit load by visual inspection has high uncertainty and is prone to errors; 3) Extrapolation of the results obtained by inspecting individual trees to estimates corresponding to entire rows or blocks relies on the grower's skill and experience; 4) Inspection of a small number of trees does not provide information about the spatial variability in the orchard; and 5) Experienced and skilled personnel to perform the sampling during a fast-moving spring may be limited or not available at the desired time. For these reasons, we pursue an automated method for estimating bloom counts from digital images.

Previous work on autonomous bloom estimation. Some previous works address the problem of autonomously assessing bloom counts using customized data acquisition setups such as structured light and multispectral imaging. Nielsen et al. (2012) use a structured light system to detect and reconstruct flower positions in peach, while Wouters et al. (2015) use multispectral imaging and detect pear flower buds; both of these methods only operate at night. Horton et al. 2016 described a system using an unmanned aerial system (UAS) or drone for peach bloom intensity estimation. Based on the premise that the photosynthetic activity of this species increases during bloom period, the system relies on multispectral aerial images of the orchard, yielding an average detection rate of 84.3% for 20 test images. This method also has the intrinsic limitation of being sensitive to illumination conditions because of its use of hard thresholds.

Other research groups used color cameras as sensors to detect bloom intensity. Underwood et al. (2016) detected flower pixels on almond trees and reported that simple thresholding operations in hue-saturation-value (HSV) space were sufficient to extract the flower pixels. However, apple trees pose a unique problem, as unlike in almond trees, leaves and flowers appear at the same time, which renders the task much more challenging. Hočevar et al. (2014) also used hard color thresholding and size features for flower localization in apple trees, such that parameters must also be adjusted whenever changes in illumination, flowering density (high/low concentration), or camera position (near/far trees) occur. Aggelopoulou et al. (2011) acquired color images in broad daylight for bloom intensity estimation, but only 113 out of 250 of which were usable due to oversaturation and distortion. The authors calibrated a linear relationship between the number of white pixels and yield. Considering the environmental and management factors that influence fruit development after the flowering stage, the usefulness of such a relationship would require further

investigation. More recently, Krikeb et al. (2017) were able to quantify blooming intensity of individual apple trees using color images, but results were season-specific and required local calibration.

Prior work by the team. The PI and co-PI recently devised novel methods for fruit flower detection in individual images, partially supported by SHAP funds. The initial approach detected apple flowers using region proposals generated by grouping similar pixels into clusters known as superpixels (Achanta et al., 2012). The region proposals were converted into feature vectors using a convolutional neural network (CNN) fined-tuned for saliency detection (Zhao et al., 2015) and with labeled flower images. In Dias et al. (2018a), we concluded that this method performed well but the superpixels were often the source of errors.

To deal with the superpixel errors, our next effort used an end-to-end residual CNN (Dias et al. 2018b). These networks produce coarse segmentations, so a refinement method allowed pixel-level segmentation of flowers (Dias and Medeiros, 2018). Our algorithm was trained using a dataset of images of apple trees collected using a hand-held camera in a USDA-ARS-AFRS orchard in Kearneysville, WV under natural daylight illumination. A total of 147 images with resolution of 5184 by 3456 pixels were acquired under multiple angles and distances. The dataset was manually labeled by indicating which pixels in each image are part of a flower using an annotation tool developed by the team (Dias et al. 2019). The images were randomly split into training and validation sets consisting of 100 and 47 images respectively. See Figure 1 for an example of this dataset and some of the results from our method. The annotated dataset was released to the public (Dias et al 2018c).

This most recent work, Dias et al. (2018b), also allows for transfer learning of a model trained on apple flowers, so peach and pear flowers can be detected as well with the same model. We evaluated the generalization capability of our method on three additional datasets. One of the additional datasets also corresponds to apple trees but with a blue background panel positioned behind the trees to visually separate them from other rows of the orchard. The second dataset contains images from a peach orchard during a cloudy day. The third dataset consists of images from a pear orchard also on a cloudy day. Our experiments demonstrated that our approach shows an optimal F1 score (i.e., a combination of recall and precision rates) of 77%, 74%, and 86% respectively for these test datasets, compared to 83% in the original test set. It should be noted that these scores are computed according very stringent pixel-wise metrics that penalize a detection for each incorrect pixel.

Figure 1. Example images, with result overlaid, from datasets acquired by consumer-level cameras. The result of running our algorithm is represented by colored outlines around regions: blue indicates a correct identification of a flower region, red indicates regions incorrectly detected as flowers, and magenta indicates non-flower regions identified as flowers.

Proposed approach. Visually separating clusters of flowers such as those seen in Figure 2 (left) is challenging. However, by observing the same set of flowers from the perspective on the right side of the figure, the problem becomes trivial. Hence, we propose to augment our low-cost sensing system for bloom estimation using consumer-grade color cameras such as those found in smartphones, hand-held digital cameras, or GoPros with a multi-view feature that associates clusters of flowers observed in one view with the corresponding separated flowers in the second view to determine the correct number of flowers

within that cluster. Classification of bloom versus non-bloom regions will be accomplished by a CNN-based method, also referred to as deep learning. Flower association will be carried out by analyzing the geometry of both scenes and aligning matching detections.

Figure 2: Two views of the same group of flowers. Left) Apparent cluster of flowers caused by self-occlusion. Right) The same flowers can be easily identified from a different perspective. We propose to use deep learning to align the left and right images and to disambiguate the incorrect count on the left image (i.e., one flower) and the correct count on the right (i.e., five flowers).

Our proposal fits in the **Ag Engineering** area and covers all of the subject areas: *Sensor Technologies*, *Use of New Technology to Improve Data Collection for Decision Making*, and *Mechanization/ Automation Technologies for Improved Farm Efficiency.* In terms of *Sensor Technologies*, we will use digital cameras for data acquisition and develop new algorithms that use sensor data as input. For the *Use of New Technology to Improve Data Collection for Decision Making*, one of the impacts of this work will be the ability to use an automated method to acquire bloom counts, which is needed for many aspects of decision making for crop management: informing bloom thinning sprays, carbohydrate models, and pollen tube models. Finally, in the *Mechanization/ Automation Technologies for Improved Farm Efficiency* the bloom estimation will be done autonomously by computer vision systems and algorithms instead of by humans, leading to increased farm efficiency.

Objectives:

1. Strategy to associate matching flower detections: investigate the applicability of end-to-end semantic alignment deep neural networks to the problem of associating flowers detected in two images containing the same tree.
2. Strategy to disambiguate flower counts: investigate the applicability of leaf counting deep neural networks to the problem of counting the number of flowers in corresponding regions of aligned images.

Procedures:

Objective 1. Traditional systems used to align two or more images (e.g., to generate photographic panoramas) are fundamentally dependent upon the fact that neighboring images share visually identical regions. In the case of flower detection, as illustrated in Figure 2, the appearance of the flowers may vary dramatically, making it impossible to apply these traditional methods. Very recently, deep neural networks have been successfully used to perform image matching in scenarios where the images are not visually similar but have semantic similarities. That is, the images are aligned not based on the appearance of the objects present in it but on their categories (e.g., two images of a car can be aligned by matching parts such as its wheels). Semantic alignment networks identify the objects of interest using detectors pre-trained on tens of thousands of images containing tens to hundreds of object categories (Caesar et al. 2016). To make the network respond to flowers, we propose to modify existing semantic alignment methods such as (Rocco et al. 2018) to replace its generic object detection module with the species-independent flower detection mechanisms that we have developed in our previous work. To generate sufficient annotated data to train the CNN, semantic alignment methods resort to synthetic data for a pre-training step, which is then followed

by a fine-tuning step using real-world data. We propose to employ a similar strategy using simulated data of flowering fruit trees, where the flower, tree, and background locations are known and consequently, annotated. This will be accomplished using L-systems, or Lindenmayer systems, a context-free grammar that has been used for representing plants (Prusinkiewicz and Lindenmayer, 1990). Fruit tree specific models and flowers can also be created with this method (Costes et al. 2008, Allen et al. 2004, Owens 2016). Suitable backgrounds are grass and other rows and can be acquired from real orchard settings. Once 3D tree models are generated, and suitable background images acquired, training data will be created by simulating cameras in OpenGL and acquiring images of the 3D scene of tree models with the background image behind it. We will then fine-tune the entire network using training dataset described above. With this approach, we expect the CNN to learn geometric transformations optimized to images acquired at orchards.

Objective 2. After two images have been aligned, if there is a disagreement regarding the number of flowers detected in a certain region, a mechanism is necessary to perform the disambiguation. Since it is not possible to determine whether multiple detections in one image correspond to a cluster of flowers in another image or if they are simply mistakes introduced by incorrect detections, we will evaluate the possibility of employing a CNN-based regression algorithm to count the number of flowers in corresponding image patches. Very recently, CNNs have been successfully employed to address the problem of leaf counting in rosette-shaped plants (Giuffirda et al. 2018, Ubbens et a. 2018). Drawing inspiration from these methods, we will integrate our flower segmentation CNNs with a regression layer that outputs the number of flowers in one image patch. However, CNN-based leaf counting methods require images of a single plant with most leaves relatively unoccluded. To account for the need to find a consensus between two image patches where only one patch is free from occlusions, we intend to further extend that approach using a Siamese structure in which two aligned image patches are simultaneously provided to the CNN. This approach will encourage the network to focus its attention on the correlated portions of the patches. Referring back to Figure 2, if the network were provided only the region on the left image, it would be difficult for it to determine that it corresponds to a cluster since the size of individual flowers varies significantly depending on the distance to the camera. By providing the network the outlined regions on both left and right images, it should be able to easily determine that the region on the left is a cluster consisting of five flowers. As in Objective 1, the network will be pre-trained using synthetic data and fine-tuned with our training set on pairs of image patches aligned using the method described in Objective 1.

Evaluation plan:
The proposed method will be evaluated on simulated data where all the elements of the scene geometry are known. Different camera displacements can be tested to evaluate the limits of how far apart the two views can be to get good flower recognitions. The system will then be evaluated on our annotated datasets of real trees in bloom, which contain a total of 209 images of apple, peach, and pear flowers, in which 7,037 flowers were manually annotated. The accuracy will be assessed according to traditional machine learning metrics such as precision-recall curves, which summarize the trade-offs between the number of incorrect detections and the number of missed detections generated by an algorithm as its decision threshold varies. Additionally, we will integrate the method with our mobile phone application that is being developed to validate the algorithms in the field and to collect additional data for training and testing our methods.

Budget:

Salaries $20,000	Supplies _____
Hourly wages _____	Travel _____
Fringe Benefits _____	Miscellaneous $200
Total $20,200	

Other Support:

Current support USDA-ARS agreement with Marquette University #58-8080-5-020, $75,875.

Current support: State Horticultural Association of Pennsylvania Research Committee, Bloom intensity estimation using your smartphone: Machine learning algorithms for species-independent visual recognition of flowers. PIs H. Medeiros and A. Tabb. $20,200.

References:

1. Achanta, R., Shaji, A., et al. (2012) SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11): 2274–2281.

2. Aggelopoulou, A., Bochtis, et al. (2011). Yield prediction in apple orchards based on image processing. Precision Agriculture 12: 448-456.

3. Allen, M., Prusinkiewicz, P. and DeJong, T. (2004) Using L-Systems for Modeling the Architecture and Physiology of Growing Trees: The L-PEACH Model. In 4th International Workshop on Functional-Structural Plant Models, pp. 220-225.

4. Auxt Baugher, T., J. Schupp, et al. (2010). String Blossom Thinner Designed for Variable Tree Forms Increases Crop Load Management Efficiency in Trials in Four United States Peach-growing Regions. Hort. Technology 20(2): 409-414.

5. Caesar, H., Uijlings, J., & Ferrari, V. (2016). COCO-Stuff: Thing and Stuff Classes in Context. arXiv preprint arXiv:1612.03716.

6. Costes, E., Smith, C., et al. (2008). MAppleT: simulation of apple tree development using mixed stochastic and biomechanical models. Functional Plant Biology 35(10), pp. 936-950.

7. Dias, P., A. Tabb, H. Medeiros. (2018). Apple Flower Detection using Deep Convolutional Networks. Computers and Industry vol. 99, pp. 17–28, Aug. 2018. doi: 10.1016/j.compind.2018.03.010.

8. Dias, P.A., A. Tabb, H. Medeiros. (2018). Multispecies fruit flower detection using a refined semantic segmentation network. Robotics and Automation Letters, vol. 3, no. 4, pp. 3003-3010, Oct. 2018.

9. Dias, P.A., Tabb, A., Medeiros, H. (2018). Data from: Multi-species fruit flower detection using a refined semantic segmentation network. Ag Data Commons. http://dx.doi.org/10.15482/USDA.ADC/1423466

10. Dias, P. A.; Medeiros, H. (2019). Semantic Segmentation Refinement by Monte Carlo Region Growing of High Confidence Detections. Asian Conference on Computer Vision (ACCV). [presented, to appear.]

11. Hočevar, M., B. Širok, et al. (2014). Flowering estimation in apple orchards by image analysis. Precision Agriculture 15 (4): 466–478. doi:10.1007/s11119-013-9341-6.

12. Horton, R., E. Cano, et al. (2016). Peach Flower Monitoring Using Aerial Multispectral Imaging, 2016 ASABE International Meeting. doi:10.13031/aim.20162461520.

13. Krikeb, O., Alchanatis, et al. (2017). Evaluation of apple flowering intensity using color image processing for tree specific chemical thinning. Advances in Animal Biosciences 8: 466-470.

14. Lyons, D.J., P.H. Heinemann, J.R. Schupp, T.A. Baugher, and J. Liu. (2015). Development of a selective automated blossom thinning system for peaches. Transactions of the ASABE 58(6):1447-1457.

15. Nielsen, M., D. C. Slaughter and C. Gliever. (2012). Vision-Based 3D Peach Tree Reconstruction for Automated Blossom Thinning*, IEEE Transactions on Industrial Informatics,* 8(1): 188-196.

16. Owens, A. (2016) Modeling Dense Inflorescences. M.Sc. thesis, University of Calgary.

17. Prusinkiewicz, P. and Lindenmayer, A. (1990). The Algorithmic Beauty of Plants. Springer-Verlag. pp. 101–107. ISBN 978-0-387-97297-8.

18. Robinson, T. S. Hoying, et al. (2014). Precision Crop Load Management, Part 2. New York Fruit Quarterly. 22(1):9-13.

19. Wouters, N., Ketelaere, B.D., et al. (2015) Multispectral detection of floral buds for automated thinning of pear, In Computers and Electronics in Agriculture, 113: 93-103.

20. Underwood, J. P., Hung, C., et al. (2016). Mapping almond orchard canopy volume, flowers, fruit and yield using LiDAR and vision sensors. Comp. and Elec in Agriculture, 130: 83-96.

21. Yoder, K.S., G.M. Peck, et al. (2012). Using a pollen tube growth model to improve apple bloom thinning for organic production. II International Organic Fruit Symposium 1001. 207-2014.

22. Zhao, R., Ouyang, W.; Li, H.; Wang, X. (2015). Saliency Detection by Multi-Context Deep Learning. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1265-1274.