An Overview of Perception Methods for Horticultural Robots: From Pollination to Harvest

Ho Seok Ahn^{1*}, Feras Dayoub², Marija Popović³, Bruce A. MacDonald¹, Roland Siegwart³, and Inkyu Sa³

Abstract—Horticultural enterprises are becoming more sophisticated as the range of the crops they target expands. Requirements for enhanced efficiency and productivity have driven the demand for automating on-field operations. However, various problems remain yet to be solved for their reliable, safe deployment in real-world scenarios. This paper examines major research trends and current challenges in horticultural robotics. Specifically, our work focuses on sensing and perception in the three main horticultural procedures: pollination, yield estimation, and harvesting. For each task, we expose major issues arising from the unstructured, cluttered, and rugged nature of field environments, including variable lighting conditions and difficulties in fruit-specific detection, and highlight promising contemporary studies. Based on our survey, we address new perspectives and discuss future development trends.

I. Introduction

The horticultural industry faces increasing pressure as demands for high-quality food, low-cost production, and environmental sustainability grow. To cater for these requirements, a top priority in this field is optimizing methods applied at each stage of the horticultural process. However, on-field procedures still rely heavily on manual tasks, which are arduous and expensive. In the past few decades, robotic technologies have emerged as a flexible, cost-efficient alternative for overcoming these issues [1].

To fully exploit the potential of automated field production techniques, several challenges remain to be addressed. Robots in orchard environments must tackle issues such as the presence of uncontrolled plant growth, weather exposure, as well as the slope, softness, cluttered and undulating nature of the transversed terrain. In addition, there are perceptual difficulties in cluttered environments due to variable illumination conditions and occlusions, as depicted in Fig. 1.

This paper summarizes recent developments and research in robotics for horticultural applications. We structure our discussion based on three main procedures in the general horticultural process, which correspond to the key stages of plant growth: (1) pollination, (2) yield estimation, and (3) harvesting. Specifically, this paper focuses on robotic

Ho Seok Ahn¹ and Bruce A. MacDonald¹ are with the Department of Electrical and Computer Engineering, CARES, University of Auckland, 368 Khyber Pass Rd Newmarket Auckland 1023, New Zealand. {hs.ahn,b.macdonald}@auckland.ac.nz

Feras Dayoub² is with the Australian Centre for Robotic Vision at Queensland University of Technology (QUT). feras.dayoub@qut.edu.au

Marija Popović³, Roland Siegwart³, and Inkyu Sa³ are with Department of Mechanical and Process Engineering, Institute of Robotics and Intelligent Systems, Autonomous Systems Lab., ETH Zurich, Switzerland. {mpopovic, rsiegwart, sain}@ethz.ch

*The corresponding author.

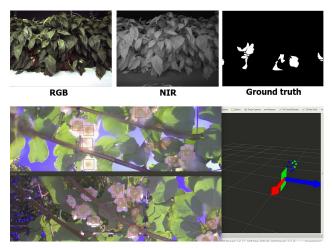


Fig. 1: Challenges in perception for horticultural applications. (Top) Red and green sweet peppers are difficult to identify in occluded conditions, even for humans. (Bottom) Matching targets detected in two camera images and finding their 3D positions is hard when perception aliasing occurs.

perception methods, which are a core requirement for systems operating in all three areas, and an actively researched field of study. For each step, we examine major challenges and outline potential areas for future work. The following provides a brief overview to motivate our study.

Pollination: Bees are major pollinators in traditional horticulture [2]. However, their numbers are diminishing rapidly due to colony collapse disorder, pesticides, and invasive mites, as well as climate change and inconsistent hive performance [3]. Studies conducted between 2015 and 2016 report a total annual colony loss of 40.5% [4] in the United States. This leads to a decrease in crop quality and quantity, causing farm owners to hire employees for seasonal hand pollination [5], which is labor-intensive. To address these issues, robotic systems are being developed to spray pollen on flowers [6]. An important consideration is producing fruit with uniform size and quality to raise their value.

Yield estimation: Crop production estimates provide valuable information for cultivators to plan and allocate resources for harvest and post-harvest activities. Currently, this process is performed by manual visual inspection of a sparse sample of crops, which is not only labour-intensive and expensive, but also inaccurate, depending on the number of counts taken, and with sometimes destructive outcomes. For this task, automated solutions have also been proposed as an alternative [7]. Here, a key aspect is designing systems able to operate in unstructured and cluttered environments [1].





Fig. 2: The robotics platform used in a mango orchard (left) [9] and a customized high-resolution stereo sensor used for imaging Sorghum stalks (right) [10].

Harvesting: The final horticultural procedure performed on-field, harvesting, usually incurs high labor cost due to its repetitive and monotonous nature. Autonomous harvesters have been proposed as a viable replacement which can also procure relatively high-quality products [8]. However, deployment in real orchards requires complex vision techniques able to handle a wide range of perceptual conditions.

This paper is organized as follows. In Section II, we examine the main sensing challenges in horticultural environments. Section III discusses flower detection and recognition methods for pollination, while Sections IV and V discuss perception challenges for automated yield estimation and harvesting, respectively. Concluding remarks including directions for further research are given in Section VI.

II. SENSING CHALLENGES IN HORTICULTURAL ENVIRONMENTS

Our survey focuses on two contemporary technologies as methods of addressing major challenges in developing robots for pollination and harvesting: sensing and perception. Successfully detecting crops and their parts plays a crucial role in horticulture, as these processes lay the groundwork for subsequent operations, such as selective spraying or weeding, obstacle avoidance, and crop picking and placing. Recent developments have enabled harvesting and scouting robots to deploy lighter, less power-demanding, but higher-resolution and faster sensors. These allow for perceiving the finer details of objects, resulting in improved performance. In the following, we elaborate on high-resolution, multi-spectral, and 3D sensing devices used for horticultural robots.

A. High-resolution sensing

Stein et al. [9] exemplified mango detection and counting by exploiting 8.14M pixel images and a Light Detection And Ranging (LiDAR) system. High-resolution images enable detecting the details of plants, which leads to the extraction of useful and distinguishable features for fruit detection. A LiDAR generates an image mask of a tree by projecting 3D points of a segmented tree back to the camera plane for associating the fruit with the corresponding tree. This study reports impressive results in fruit counting (1.4% over counting), and precision and accuracy ($R^2 = 0.90$).

High-resolution cameras are not only useful for crop field scouting missions from the air, but also on the ground. Beweja et al. [10] used a 9M pixel stereo-camera pair with a high-power flash triggered at 3 Hz (Fig. 2). Stalks of Sorghum plants and their widths were detected and estimated for subsequent phenotyping. A disparity map calculated from the stereo images enables metric measurements with high precision (mean absolute error of 2.77 mm). They report their approach to be 30 and 270 times faster for counting and width measurements respectively compared to conventional manual methods. However, only a limited number of stalks were considered for counting (24) and width estimation (17).

B. Multi-spectral sensing

Detection performance can be improved by observing the infrared (IR) range in addition to the visible spectrum through multi-spectral (RGB+IR) images. In the past, this sensor was very costly due to the laborious manufacturing procedure behind multi-channel imagers, but they are now more affordable, with off-the-shelf commercial products readily available thanks to advances in sensing technologies.

There are substantial studies on using multi-spectral cameras on harvesting and scouting robots in orchards [11] and open fields (sweet peppers) [12], [13]. The use of multi-spectral images improves classification performance by about 10% with respect to RGB only models, with a global accuracy of 88%. This improvement is analogous to that in [12], [13] for detecting sweet peppers.

C. From 2D to 3D: RGB-D sensing and LiDAR

Thus far, our work discussed the usage of 2D and passive sensing technologies for harvesting robots. However, advances in integrated circuit and microelectromechanical system (MEMs) technologies also unlock the potential of 3D sensing devices. For example, Red, Green, Blue, and Depth (RGB-D) sensing allows for constructing metric maps with high accuracy, as shown in Fig. 3. This information is useful, not only for object detection, classification [14], and fruit localization, but also for motion planning [15], obstacle avoidance, and fine end-effector operation. A crucial step in the operation of RGB-D sensors is filtering out noise (denoising) and outliers, which may be caused by poor sensor calibration, inaccurate disparity measurements due to illreflectance (e.g., under direct sunlight), etc., as highlighted in [16]. Using LiDAR is beneficial for longer-range scanning and mapping large fields [17], [18] (Fig. 3). It is also necessary to design an RGB-D or LiDAR based harvesting system that can handle large amounts of incoming 3D data, which influences the cycle time of a harvesting robot, or the time for picking and placing a fruit. Table I summarizes our review of sensing technologies for harvesting robots.

III. PERCEPTION FOR POLLINATION

Early flower detection methods, e.g., in [19], rely mainly on color values and do not perform well as many flowers have similar colors. To solve this issue, recent works consider additional information such as size, shape, and feature edges. Nilsback et al. [20] used color and shape. Pornpanomchai et al. [21] used RGB values with the flower size and the edge of petals feature to find herb flowers. Hong et al.

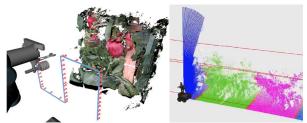


Fig. 3: (Left) RGB-D sensing used for a protective sweet pepper farm reconstruction [15]. (Right) LiDAR mapping of almond trees in an orchard [18].

TABLE I: Sensors for harvesting robots

Sensors	Addressed challenges	Crons	Sensor spec.	Paper	Acc.a
High-res	Deficiency	Mango	8.14MP	[9]	90 ^a
camera	in detail	Sorghum	9MP	[10]	0.88^{a}
Multi-spectral	Improve	Almond	RGB+IR	[11]	88%
camera	accuracy	Sweet pepper	RGB+IR	[12], [13]	69.2%,58.9%
	Metric,	Sweet pepper	RGB-D	[14], [15]	80%, 0.71°
3D sensing	motion	Almond trees	LiDAR	[18]	0.77^{a}
	planning				

^a R-squared correlation. ^b Picking successful rate. ^c AUC of peduncle detection rate.

[22] found the contour of a flower image using both colorand edge-based contour detection. Tiay et al. [23] used a similar method, which uses edge and color characteristics of flower images. Yuan et al. [24] used hue, saturation, and intensity (HSI) values to find flower, leaf and stem areas, before reducing noise with median and size filters. Bosch et al. [25] proposed image classification using random forests and compared multi-way Support Vector Machines (SVMs) with region of interest, visual word, and edge distributions. Kaur et al. [26] identified rose flowers using the Otsu algorithm and morphological operations. Bairwa et al. [27] and Abinaya et al. [28] proposed thresholding techniques to count gerbera and jasmine flowers, respectively. However, these color based approaches are not robust enough in variable lighting conditions.

To address this, recent research has considered deep learning techniques for flower detection. Yahata et al. [29] proposed a hybrid image sensing method for flower phenotyping. Their pipeline is based on a coarse-to-fine approach, where candidate flower regions are first detected by Simple Linear Iterative Clustering (SLIC) and hue channel information, before the acceptance of flowers is decided by a convolutional neural network (CNN). Liu et al. [30] also developed a flower detection method based on CNNs. Srinivasan et al. [31] developed a machine learning algorithm that receives a 2D RGB image and synthesizes an RGB-D light field (scene color and depth in each ray direction). It consists of a CNN that estimates scene geometry, a stage that renders a Lambertian light field using that geometry, and a second CNN that predicts occluded rays and non-Lambertian effects. While these approaches perform better compared to traditional methods, they demand high training times with large datasets on high-performance systems.

IV. PERCEPTION FOR YIELD ESTIMATION

Manual yield estimation is time-consuming, expensive, labor-intensive and inaccurate, with sometimes destructive outcomes. These aspects have motivated methods of process automation. However, fully automated estimation is a challenging task as: (a) the environment is unstructured and cluttered, (b) the fruit can have colors similar to the background, (c) they may lack distinguishable features and be occluded by other fruit, branches, or leaves, and above all, (d) there are uncontrolled illumination changes when yield estimation is done outdoors. In the following subsections, we elaborate on two different paradigms tackling these issues.

A. Hand-crafted features

Traditional yield estimation algorithms rely on visual detection methods using predefined hand-crafted features derived from image content. These can be based on various information, such as the shape, color, texture, or spatial orientation of the fruit using various feature representations such as the local binary pattern (LBP, texture features) [32] and a histogram of gradients (HoG, geometry and structure features) [33], SIFT [34] or SURF [35] (key-points features). For example, Nuske et al. [7] and Li et al. [36] utilized shape and texture features to detect grapes and green apples, respectively. Wang et al. [37] and Linker [38] exploited color and specular reflection to detect apples. Verma et al. [39] and Dorj et al. [40] used color-space features to detect tomatoes and tangerines, respectively.

The recent survey by Gongal et al. [41] reviews computer vision for fruit detection and localization. This paper draws the following conclusions:

- learning-based methods are superior to simple threshold-based image segmentation methods for fruit detection in realistic environments,
- combining multiple types of hand-crafted features is better than using only one type of feature,
- detection methods based on hand-crafted features perform poorly when faced with occlusions, overlapping fruit and variable lighting conditions.

After the above survey was published, there was a major breakthrough in object detection and localization using Deep Neural Networks (DNNs) for learned feature extraction. In the next section, we examine the new shift towards deep learning for yield estimation applications.

B. Learning-based features

One of the earliest works using learned features was applied to segment almond fruit [11] (Fig. 4). Visual features from multi-spectral images were learned using a sparse autoencoder at different image scales, followed by a logistic regression classifier to learn pixel label associations.

Results show that leveraging a learning approach for feature extraction renders the system more robust to illumination changes. Stein et al. [42] proposed a mango fruit detection and tracking system based on a faster region-based Convolutional Neural Network (Faster R-CNN) [43] using

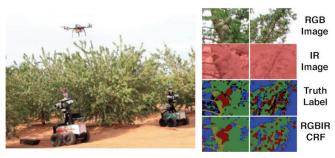


Fig. 4: Orchard almond yield estimation using multi-spectral images [11].

detection and camera trajectory information to establish pairwise correspondences between consecutive images.

Rahnemoonfar and Sheppard [44] proposed a fruit counting system that employs a modified version of the Inception-ResNet architecture [45] trained on synthetic data. The network predicts the number of fruit from the input image directly, without segmentation or object detection as intermediate steps. Recently, Halstead et al. [46] proposed a sweet pepper detection and tracking system inspired by the DeepFruits detector [47]. The system is trained to perform efficient in-field assessment of both fruit quantity and quality.

Despite the high accuracy measures reported in the works above, using deep learning for fruit detection still faces the following challenges: (a) It requires vast amounts of labeled data, which is time-consuming and can be expensive to obtain. Using synthetic data from generative models, e.g., in [44] and [48], has recently emerged as a method of addressing this issue. (b) Tracking and data association is crucial to prevent over-counting produce. Not all works reviewed address this aspect, as they estimate yield from only a single image of a tree such that only fruit in the current view are counted. Manual calibration is usually carried out to infer the total amount of fruit based on the visible proportion. However, this process is required for every tree species, and also varies between years. (c) There is a lack of independent third-party benchmark tests for estimating yields for various fruit. Currently, results are reported based on inhouse datasets, which can be very small, making reported results hard to compare fairly.

V. PERCEPTION FOR HARVESTING

Traditionally, harvesting robots exploit traditional hand-crafted approaches (e.g., extracting useful visual or geometric features) [13], [14] and plantation geometries (e.g., tree rows in orchards) [18] for crop perception. While these methods show promising results, our survey concentrates on the new paradigm of using data-driven DNNs. The variety of off-the-shelf DNNs today with human-level performance [49], [50] indicates they are transitioning from dataset benchmarking into in-situ production environments. In this section, we investigate strategies for tackling the major perceptual challenges in automated harvesting: occlusions, perception aliasing, and environmental variability.



Fig. 5: Examples of detecting (top) strawberry and (bottom) mango fruits using the DeepFruits network [47].

A. Occlusions

Harvesting scenes commonly exhibit occlusions of crops caused by themselves or other plant parts (leafs, stems or peduncles). As a result, it may be difficult to detect crops using only low-level features, such as colour, texture, and shape, etc., and instead necessary to employ DNNs for higher-level contextual scene understanding through convolutional multi layers. To this end, typically a large number of internal parameters (e.g., weights and biases) must be properly tuned, which is a nontrivial task. This training process requires a vast number of samples to avoid over-fitting to small datasets. It is thus common practice to use pre-trained parameters fitted on millions of samples (e.g., images) for variable initialization; to fine-tune. After fine-tuning, the trained DNN can be refined with relatively small datasets. Sa et al. [47] exemplified this procedure, as shown in Fig. 5. 602 images of seven fruit were considered for model training and testing with ~0.9 F1-score for most fruit. Although the trained model handles occlusions reasonably, it struggles with large variations in training and testing images as it expects visually similar environments.

B. Homogeneous farm fields: perception aliasing

It is common practice to seed crops and trees in linear row arrangements separated by vacant space. This geometric formation can be useful for perception by providing an informed prior for plant detection [51]. However, such a structure also creates the issue of perception aliasing, which particularly impedes accurate mapping and localization in farm fields. Many crops resembling each other in appearance produce high false positive rates, which degrades harvesting performance. Recently, Kraemer et al. [52] proposed a method of creating landmarks from plants using an FCNN. These landmarks can be used for robot pose localization and mapping farm environments. It is also possible to address this issue by fusing different sensing modalities, e.g., from an Inertial Measurement Unit (IMU), wheel odometry, LiDAR, or Global Positioning System (GPS), to bound a search space for visual matching.

TABLE II: Perception trend of harvesting robots

Paper	Network arch.	Crops	Classification type	# images (tr/test)	Accu.
[47]	Faster R-CNN	7 fruits	Obj. based	484/118 RGB	0.9 ^a
[52]	FCNN	Root crops	Pixel based	1398/300 RGB+IR	0.9 ^b
[53]	FCNN	Orchard fruits	Pixel and obj. based	47/45 RGB	0.838 ^c
[54]	Segnet	Root crops	Pixel based	375/90 RGB+IR	0.8ª

^a F1-score ^b precision/recall rate

C. Environmental variability

Harvesting robots usually operate in a wide range of perceptual conditions, which may feature various lightning conditions, dynamic objects, and unbounded crop scales. Researchers have attempted to restrict operating environments using light-controllable greenhouses or by operating at night time with high-power artificial flashes rather than harvesting in open fields. While these constraints increase production costs and reduce operation time, they do improve perceptual performance. Chen et al. [53] demonstrated an approach to detect and count oranges and apples with data-driven FCNN. A mean intersection of union of 0.813 and 0.838 were achieved for oranges and apples, respectively. Table II summarizes our study of perception for harvesting robots.

VI. CONCLUSIONS

To the best of the authors' knowledge, this survey is the first of its kind to cope with the three major inter-connected horticultural procedures of pollination, yield estimation, and harvesting in the context of autonomous robotics. Our discussions in this work target practical challenges in perception that inhibit robot deployment in real production scenarios. We tried to supplement, rather than replicate, current survey papers by focusing on most contemporary works not covered by these reviews.

Our survey exposed that the main challenges facing the development of automated pollination robots involve selecting hardware and sensing equipment, robust flower detection, and row-following on uneven and bumpy surfaces. Whereas traditional hand-crafted features with classifiers have been widely exploited for flower detection, there is a rapid paradigm shift towards DNNs. Complementary multimodal sensing is an essential element for robust vehicle navigation to compensate for uneven outdoor environments.

In automated yield estimation, our survey revealed challenges in crop detection in unstructured and cluttered environments, uncontrolled lighting conditions, and occlusions caused by leaves, branches, and other crops. Here, machine learning also plays a pivotal role in crop detection and localization, and DNNs pave the way for overcoming occlusions and illumination changes.

Issues in automated harvesting closely resemble those in yield estimation, with additional difficulties arising in designing manipulators and end-effectors. Unless there are special requirements, it is desirable to use commercial manipulators to minimize development time and effort, with only a custom and application-specific end-effector design. Exploiting geometrical prior knowledge about fields, such as crop rows, to improve performance is viable for common horticultural practices.

While our review of recent developments in DNNs and GPU-driven computing uncovers their potential in horticulture, several open challenges remain. Namely, the state-of-art requires larger, more accessible datasets to prevent cases of model over-fitting, as well as faster processing devices to enable real field deployments.

ACKNOWLEDGEMENT

This study has received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644227, (Flourish) from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0029. It was also supported by the New Zealand Ministry for Business, Innovation and Employment (MBIE) on contract UOAX1414. We would like to thank to New Zealand MBIE Orchard project team (The University of Auckland, Plant and Food Research, The University of Waikato, Robotics Plus).

REFERENCES

- [1] A. Bechar and C. Vigneault, "Agricultural robots for field operations. part 2: Operations and systems," *Biosystems Engineering*, vol. 153, pp. 110 – 128, 2017.
- [2] D. M. Lofaro, "The honey bee initiative smart hive," in *International Conference on Ubiquitous Robots and Ambient Intelligence*, pp. 446–447, 2017.
- [3] D. Goulson, E. Nicholls, C. Botías, and E. L. Rotheray, "Bee declines driven by combined stress from parasites, pesticides, and lack of flowers," *Science*, vol. 347, no. 6229, 2015.
- [4] K. Kulhanek, N. Steinhauer, K. Rennich, D. M. Caron, R. R. Sagili, J. S. Pettis, J. D. Ellis, M. E. Wilson, J. T. Wilkes, D. R. Tarpy, R. Rose, K. Lee, J. Rangel, and D. vanEngelsdorp, "A national survey of managed honey bee 2015–2016 annual colony losses in the usa," *Journal of Apicultural Research*, vol. 56, no. 4, pp. 328–340, 2017.
- [5] T. Giang, J. Kuroda, and T. Shaneyfelt, "Implementation of automated vanilla pollination robotic crane prototype," in *System of Systems Engineering Conference*, 2017.
- [6] R. N. Abutalipov, Y. V. Bolgov, and H. M. Senov, "Flowering plants pollination robotic system for greenhouses by means of nano copter (drone aircraft)," in *IEEE Conference on Quality Management, Transport and Information Security, Information Technologies*, pp. 7– 9, Oct 2016.
- [7] S. Nuske, S. Achar, T. Bates, S. G. Narasimhan, and S. Singh, "Yield estimation in vineyards by visual grape detection," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2352–2358, 2011.
- [8] C. Lehnert, A. English, C. McCool, A. W. Tow, and T. Perez, "Autonomous sweet pepper harvesting for protected cropping systems," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 872–879, 2017.
- [9] M. Stein, S. Bargoti, and J. Underwood, "Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry," Sensors, vol. 16, p. 1915, Nov. 2016.
- [10] H. S. Baweja, T. Parhar, O. Mirbod, and S. Nuske, "StalkNet: A Deep Learning Pipeline for High-Throughput Measurement of Plant Stalk Count and Stalk Width," in *Field and Service Robotics*, pp. 271–284, Springer, 2018.
- [11] C. Hung, J. Nieto, Z. Taylor, J. Underwood, and S. Sukkarieh, "Orchard fruit segmentation using multi-spectral feature learning," in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5314–5320, IEEE, 2013.
- [12] C. McCool, I. Sa, F. Dayoub, C. Lehnert, T. Perez, and B. Upcroft, "Visual detection of occluded crop: For automated harvesting," in *IEEE International Conference on Robotics and Automation*, pp. 2506–2512, IEEE, 2016.

c mean intersection of union

- [13] C. Bac, J. Hemming, and E. Van Henten, "Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper," *Computers and electronics in agriculture*, vol. 96, pp. 148–162, 2013.
- [14] I. Sa, C. Lehnert, A. English, C. McCool, F. Dayoub, B. Upcroft, and T. Perez, "Peduncle detection of sweet pepper for autonomous crop harvesting—Combined Color and 3-D Information," *IEEE Robotics* and Automation Letters, vol. 2, no. 2, pp. 765–772, 2017.
- [15] C. Lehnert, I. Sa, C. McCool, B. Upcroft, and T. Perez, "Sweet Pepper Pose Detection and Grasping for Automated Crop Harvesting," in IEEE International Conference on Robotics and Automation, 2016.
- [16] M. Firman, "RGBD datasets: Past, present and future," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 19–31, 2016.
- [17] S. Bargoti, Fruit Detection and Tree Segmentation for Yield Mapping in Orchards. PhD thesis, University of Sydney, 2017.
- [18] J. P. Underwood, C. Hung, B. Whelan, and S. Sukkarieh, "Mapping almond orchard canopy volume, flowers, fruit and yield using Li-DAR and vision sensors," *Computers and Electronics in Agriculture*, vol. 130, pp. 83–96, 2016.
- [19] M. Das, R. Manmatha, and E. M. Riseman, "Indexing flower patent images using domain knowledge," in *IEEE Intelligent Systems and their Applications*, vol. 14, pp. 24–33, 1999.
- [20] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1447–1454, 2006.
- [21] C. Pornpanomchai, P. Sakunreraratsame, R. Wongsasirinart, and N. Youngtavichavhart, "Herb flower recognition system (HFRS)," in *International Conference on Electronics and Information Engineering*, pp. 123–127, 2017.
- [22] S.-W. Hong and L. Choi, "Automatic recognition of flowers through color and edge based contour detection," in *International Conference* on *Image Processing Theory, Tools and Applications*, 2012.
- [23] T. Tiay, P. Benyaphaichit, and P. Riyamongkol, "Flower recognition system based on image processing," in *ICT International Student Project Conference*, pp. 99–102, 2014.
- [24] T. Yuan, S. Zhang, X. Sheng, D. Wang, Y. Gong, and W. Li, "An autonomous pollination robot for hormone treatment of tomato flower in greenhouse," in *International Conference on Systems and Informatics*, pp. 108–113, 2016.
- [25] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *IEEE International Conference on Computer Vision*, pp. 1–8, 2007.
- [26] R. Kaur and S. Porwal, "An optimized computer vision approach to precise well-bloomed flower yielding prediction using image segmentation," in *International Journal of Computer Application*, vol. 119, pp. 15–20, 2015.
- [27] N. Bairwa and N. kumar Agrawal, "Counting of Flowers using Image Processing," in *International Journal of Engineering Research and Technology*, vol. 3, pp. 775–779, 2014.
- [28] A. Abinaya and S. M. M. Roomi, "Jasmine flower segmentation: A superpixel based approach," in *International Conference on Commu*nication and Electronics Systems, pp. 1–4, 2016.
- [29] S. Yahata, T. Onishi, K. Yamaguchiv, S. Ozawa, J. Kitazono, T. Ohkawa, T. Yoshida, N. Murakami, and H. Tsuji, "A hybrid machine learning approach to automatic plant phenotyping for smart agriculture," in *International Joint Conference on Neural Networks*, pp. 110–116, 2017.
- [30] Y. Liu, F. Tang, D. Zhou, Y. Meng, and W. Dong, "Flower classification via convolutional neural network," in *IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications*, pp. 1787–1793, 2016.
- [31] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to synthesize a 4D RGBD light field from a single image," in *IEEE International Conference on Computer Vision*, pp. 2262–2270, 2017
- [32] T. Ojala, M. Pietikainen, and others, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.
- [33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conference on Computer Vision* and Pattern Recognition, 2005.

- [34] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, Nov. 2004.
- [35] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision*, pp. 404–417, Springer Berlin Heidelberg, 2006.
- [36] D. Li, M. Shen, D. Li, and X. Yu, "Green apple recognition method based on the combination of texture and shape features," *IEEE International Conference on Mechatronics and Automation*, pp. 264–269, 2017.
- [37] Q. Wang, S. Nuske, M. Bergerman, and S. Singh, "Automated crop yield estimation for apple orchards," in *International Symposium on Experimental Robotics*, 2012.
- [38] R. Linker, "A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation," *Precision Agriculture*, vol. 18, pp. 59– 75, 2016.
- [39] U. Verma, F. Rossant, I. Bloch, J. Orensanz, and D. Boisgontier, "Segmentation of tomatoes in open field images with shape and temporal constraints," 2014.
- [40] U.-O. Dorj, K. kwang Lee, and M. Lee, "A computer vision algorithm for tangerine yield estimation," *International Journal of Bio-Science* and Bio-Technology, vol. 5, pp. 101–110, 2013.
- [41] A. Gongal, S. Amatya, M. Karkee, Q. Zhang, and K. Lewis, "Sensors and systems for fruit detection and localization: A review," *Computers* and Electronics in Agriculture, vol. 116, pp. 8–19, 2015.
- [42] M. Stein, S. Bargoti, and J. P. Underwood, "Image based mango fruit detection, localisation and yield estimation using multiple view geometry," in *Sensors*, 2016.
- [43] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137–1149, 2015.
- [44] M. Rahnemoonfar and C. Sheppard, "Deep Count: Fruit Counting Based on Deep Simulated Learning," in *Sensors*, vol. 17, 2017.
- [45] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in AAAI, 2017.
- [46] M. Halstead, C. McCool, S. Denman, T. Perez, and C. Fookes, "Fruit quantity and quality estimation using a robotic vision system," *CoRR*, vol. abs/1801.05560, 2018.
- [47] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "Deepfruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, 2016.
- [48] R. Barth, J. IJsselmuiden, J. Hemming, and E. V. Henten, "Data synthesis methods for semantic segmentation in agriculture: A capsicum annuum dataset," *Computers and Electronics in Agriculture*, vol. 144, pp. 284 296, 2018.
- [49] R. Geirhos, D. H. Janssen, H. H. Schütt, J. Rauber, M. Bethge, and F. A. Wichmann, "Comparing deep neural networks against humans: object recognition when the signal gets weaker," arXiv preprint arXiv:1706.06969, 2017.
- [50] M. P. Eckstein, K. Koehler, L. E. Welbourne, and E. Akbas, "Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes," *Current Biology*, vol. 27, no. 18, pp. 2827–2832, 2017.
- [51] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "UAV-Based Crop and Weed Classification for Smart Farming," in *IEEE International Conference on Robotics and Automation*, pp. 3024–3031, IEEE, 2017.
- [52] F. Kraemer, A. Schaefer, A. Eitel, J. Vertens, and W. Burgard, "From Plants to Landmarks: Time-invariant Plant Localization that uses Deep Pose Regression in Agricultural Fields," arXiv preprint arXiv:1709.04751, 2017.
- [53] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, C. J. Taylor, and V. Kumar, "Counting apples and oranges with deep learning: a data-driven approach," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 781–788, 2017.
- [54] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "weedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2018.