Fruit load estimation in mango orchards – a method comparison

J. P. Underwood[1], M.M. Rahman[2], A. Robson[2], K. B. Walsh[3], A. Koirala[3], Z. Wang[3]

***Abstract*—The fruit load of entire mango orchards was estimated well before harvest using (i) in-field machine vision on mobile platforms and (ii) WorldView-3 satellite imagery. For in-field machine vision, two imaging platforms were utilized, with (a) day time imaging with LiDAR based tree segmentation and multiple views per tree, and (b) night time imaging system using two images per tree. The machine vision approaches involved training of neural networks with image snips from one orchard only, followed by use for all other orchards (varying in location and cultivar). Estimates of fruit load per tree achieved up to a R^2 = 0.88 and a RMSE = 22.5 fruit/tree against harvest fruit count per tree (n = 18 trees per orchard). With satellite imaging, a regression was established between a number of spectral indices and fruit number for a set (n=18) of trees in each orchard (example: R^2 = 0.57, RMSE = 22 fruit/tree), and this model applied across all tree associated pixels per orchard. The weighted average percentage error on packhouse counts (weighted by packhouse fruit numbers) was 6.0, 8.8 and 9.9% for the day imaging system, night imaging machine vision system and the satellite method, respectively, averaged across all orchards assessed. Additionally, fruit sizing was achieved with a RMSE = 5 mm (on fruit length and width). These estimates are useful for harvest resource planning and marketing and set the foundation for automated harvest.**

I. Introduction

The management of a tree crop harvest can be aided by pre-harvest estimations of fruit load in context of harvest logistics (e.g. labour, field equipment and packing material requirements), transport and storage requirements and marketing activity (e.g. forward selling). There is potential for yield estimation of tree fruit crops using remote sensing or machine vision. Fruit load per tree could be proportional to canopy volume or foliage health, with canopy volume potentially indexed by canopy area as assessed in satellite imagery, and canopy health potentially indexed by a spectral index from the satellite multispectral data. Alternately, fruit number and size could be assessed by machine vision using imagery collected from a mobile ground based platform. At the very least, these techniques could provide a categorization of trees by level of crop load, allowing for a reduced manual sampling effort. In the current study we report on a parallel investigation of ground and satellite based estimation for mango crop load, with application across multiple orchards.

A WorldView-3 satellite imagery based system for estimation of fruit load based on canopy health and area has been previously described for macadamia and avocado by [1] (e.g. R^2 = 0.81, 0.68 and 0.72 on three avocado blocks, respectively; for 18 trees per block). This remote sensing approach allows for a yield forecast of large areas. However, Robson *et al.* [1] reported that the VI used and the relationship slope between that VI and fruit load varied with orchard, necessitating field-work for calibration at each orchard.

Machine vision systems for estimation of mango fruit load on tree have been previously reported by Payne *et al.* [2] and Stein *et al.* [3]. The former study [2] used a night imaging, dual-view approach (two images of each tree, from the two inter rows), with a report of a coefficient of determination (R^2) of only 0.74 relative to human count, with fruit number under estimated due to canopy occlusion. Stein *et al*. [3] employed a faster regional convolutional neural network (R-CNN) detector with both a dual-view and a multi-view (multiple images of each tree, from the different angles and two inter rows) approach, with the latter using trajectory data to track fruits between frames to avoid repetitive counts. Detected fruits were ascribed to individual trees by reference to a LiDAR-projected 2D image mask and 3D locations of every fruit were calculated by triangulation. The multi-view count was precise (R^2 = 0.90) and accurate (slope = 1.01) while the dual-view approach was more precise (R^2 = 0.94) but less accurate (slope = 0.54) for the assessed set of 18 sample trees. A unity slope implies a correction factor and associated field calibration work may not be required, but denser canopies still occlude some fruit, necessitating calibration even for the multi-view approach. In the current study, image training for machine learning was based on snips from one site only, with the model used on other fields. Correction factors (harvest to machine vision count) were calculated per orchard. The use of in-field machine vision to estimate fruit size was also demonstrated, extending the work of Wang *et al.* [4].

II. MATERIALS AND METHODS

A. Field material and harvest

Field work was undertaken in one orchard in the 2016/17 season and five orchards in the 2017/2018 seasons, involving different farms and mango cultivars. Orchard 1 was a commercial mango (cultivar Calypso™) block near Bundaberg, Australia. This site was also utilized in the study of Stein *et al.* [3] in a previous year. Orchard 2 was an adjacent block of the same cultivar, with larger trees. Orchard 3 was on the same farm and of the same cultivar, with yet larger trees with intertwined neighbouring tree canopies. This orchard was located about 1 km from Orchard 1. Orchards 4 and 5 were on different farms in the same district. Orchard 4 featured the R2E2 cultivar with one large completely intertwined block and

[1] J. P. Underwood is with Australian Centre for Field Robotics, University of Sydney, Sydney, Australia (e-mail: j.underwood @acfr.usyd.edu.au).

[2] M. M. Rahman and A. Robson are with Precision Agriculture Group, University of New England, Armidale, Australia.

[3] K. B. Walsh, A. Koirala and Z. Wang are with the Institute for Future Farming Systems, Central Queensland University, Rockhampton, Australia.

a second small block with well separated trees, while Orchard 5 featured relatively small, well separated trees of Honey Gold™ cultivar. Orchard 6 was located some 400 km distant, and involved large trees with intertwined canopies of the cultivar Honey Gold.

For each orchard, a set of ca.18 'calibration' trees were selected on the basis of canopy NDVI (six each of high, medium and low values; based on WV3 satellite imagery, see description in next section). Fruit load on the calibration trees of Orchard 1 was assessed by a human in-field count at the time of in-field imaging. For all orchards, the calibration trees were harvested and fruit count recorded. Each block was commercially harvested on the following day, with fruit passed over a commercial fruit grader which provided total orchard yield (# of fruit and # of trays). Fruit of consistent size were packed to trays of 7.5 kg weight, with tray sizes ranging from 12 to 28 (fruit per tray). Harvest occurred during December - January in 2017-18.

In current best farm practice, a manual count of fruit per tree is undertaken of a sample of trees to obtain an orchard yield estimate. This practice was mimicked using the calibration tree harvests to estimate total crop yield.

B. In-field machine vision imaging

Two imaging platforms were employed (Figure 1). Platform A (based on that of Stein *et al.* [3]) is based on an autonomous platform equipped with a high-precision GPS/INS device, LiDAR and a 12 MP RGB camera. The platform was operated in daylight hours, using high intensity Xe flash lamps to enable short exposure times to avoid motion blur and background light variation issues. Platform B (based on that of Payne *et al.* [2]) was mounted on a farm vehicle and was equipped with a GPS device utilized for triggering of imaging based on previously mapped tree locations, a 5 MP RGB camera, a depth camera (Microsoft Kinect v2) and 720 W of LED lighting, with imaging undertaken at night. Both platforms were operated at a speed of approximately 5 km/h, with imaging of a block of 1000 trees completed in 2 h.

Figure 1. Imaging platform A (left) and B (right).

Platform A. The techniques involved in LiDAR based estimates of canopy volume and multi-view machine vision estimates of fruit load per tree are described by Stein *et al.* [3] and Bargoti and Underwood [5]. A faster region convolutional neural network (FR-CNN) algorithm was trained on >1500 snips from images collected of a Calypso™ orchard (Orchard 1) in the previous 2016-7 season, with the trained model applied to all orchards imaged in the 2017-18 year. An example of detection is shown in Figure 2.

Fruit detection was undertaken both for a single image per side of the tree ('dual-view') and for tracked fruit in ca. 25 images per side of each tree ('multi-view'). Fruit were detected from multiple images to reduce the effect of occlusion from any single viewpoint, and were tracked to minimise redundantly counting the same fruit several times. This was achieved by associating detections from frame to frame, using GPS/INS to estimate the camera trajectory. Fruit detections within each image were associated to a specific tree by projecting segmented LiDAR data of each tree to the corresponding image frames to form image masks. To mitigate double counting of fruit apparent in the images from opposing sides of the tree, the method of Stein *et al.* [3] was modified by clipping the LiDAR masks at a vertical plane at the geometric centroid of each tree.

The performance per-tree was assessed by comparison to a manual count of fruit on each of 18 trees in Orchard 1, in December 2017 with fruit still on the tree, six weeks prior to harvest. For block-level estimates, a calibration factor was obtained from post-harvest counts of 18 trees per block from the previous year. Consequently, the reported block estimates for platform A were true predictions that were provided to farm management 6 weeks prior to harvest. The system was used as a 'black box' with no labour-intensive training, calibration or refinement using data from the current year.

Platform B. With night-time imaging, the dark background of the target tree was easily segmented using Otsu's thresholding [6] technique. This technique also enabled segmentation of tree canopies except for orchards where neighboring tree canopies were overlapped, in which case a fixed margin crop was applied. Images of trees in two rows of Orchard 1 (excluding calibration trees) were used for deep learning training, with ground truth bounding boxes. The deep learning technique of Faster R-CNN (VGG) and YOLO v3 [7] were trained with 11820 annotated fruit in 1,300 tiles, with tuning based on a further 861 fruit in 130 tiles. The trained model was then used to localize and count fruit per image for all orchards, with the fruit number per tree estimated as the sum of count for images of both tree sides ('dual-view'). However, not all fruit in the canopy can be captured in the dual-view images, so a correction factor was required. This can come from the historical yield results or from a manual fruit counts on representative sample trees. In this study the correction factor was based on the calibration tree harvest undertaken for each orchard. In addition, camera depth information enabled on-tree fruit sizing as described in [4]. A color-based thresholding was used to remove the background (e.g. leaves) from YOLO v3 output, and an elliptical fitting method was then applied to identify well separated fruit. Image size and depth information was used to estimate fruit size, which is correlated to fruit mass [8, 9].

Figure 2. Examples of imaging and fruit detection from Platform A (top) and B (bottom)

C. Satellite imaging

WV3 satellite imagery of Orchard 1 was obtained on August 18, 2017, some 5 months before harvest. The method of Robson *et al.* [1, 10] was employed. Briefly, an unsupervised classification of trees to high, medium and low Normalised Difference Vegetation Index (NDVI) categories (NDVI = (NIR1 – R)/(NIR1 + R)) was undertaken, with random selection of six trees from each class for calibration activities. A 1.5 m radius area around the central point of each calibration tree was segmented using ArcGIS 10.2 (Environmental Systems Research Institute, Redlands, CA, USA), and 18 indices relevant to crop biomass calculated for each tree [1]. Fruit number per tree was regressed against these indices. The index with the highest regression coefficient of determination for each orchard was adopted, with the model applied to the canopy associated average index value for the whole orchard. Tree canopy pixels were segmented using a 2D scatter plot (NIR1 against Red). The estimated average yield per tree for the orchard was multiplied by tree number to provide an estimate of orchard fruit yield.

III. Results and Discussion

A. Satellite image calibration

Of the 18 spectral indices calculated from the satellite imagery, the best relationship was obtained with the N2RENDVI = ((NIR2−RedEdge)/(NIR2 + RedEdge)) for orchard 1, 2 and 4; N1RENDVI = ((NIR1−RedEdge)/(NIR1 + RedEdge)) for orchard 3; and N2NDVI = ((NIR2−R)/(NIR2 + R)) for orchard 5. The R^2 of the linear regression between harvested fruit number per tree and VI ranged between 0.10 and 0.57, while RMSE varied between 22 and 151 fruit per tree across the assessed orchards (data not shown, see Figure 3 for example).

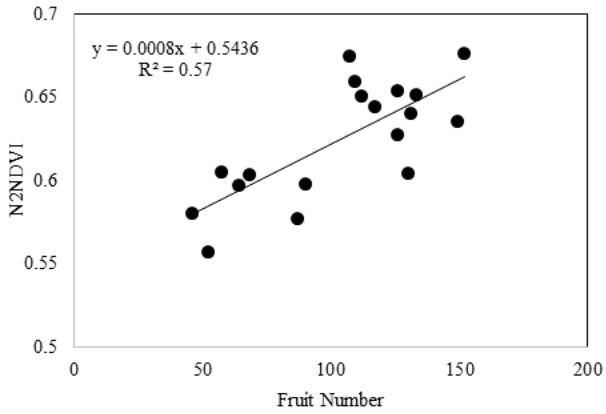


Figure 3. Example regression of satellite imaging VI on fruit load per tree.

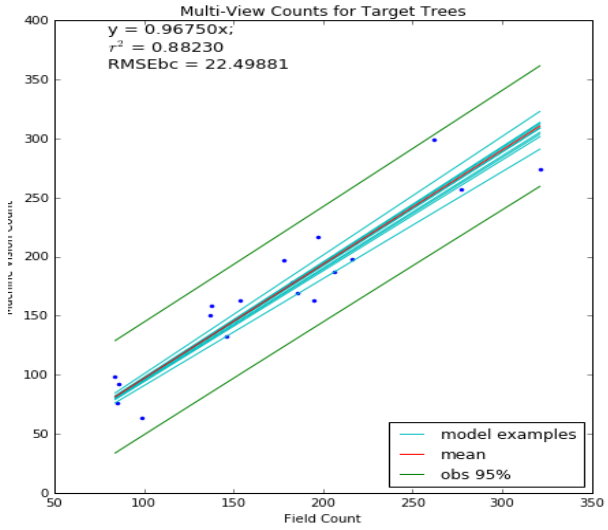
B. In-field machine vision calibration

Platform A. The FR-CNN algorithm was trained on images from the previous season. Correction factors calibrated by comparing machine vision counts to post-harvest individual tree counts from the previous year resulted in a value of 0.96 for separated Calypso trees (Orchard 1), 0.95 for a young R2E2 block (Orchard 4) and 0.875 for a mature R2E2 block (Orchard 5). The linear regression of multi-view machine vision counts to pre-harvest manual field counts in Orchard 1 in the current year was described by a similar slope of 0.97, R^2 of 0.88 and RMSE of 22 fruit (see Figure 4a), which confirmed the appropriateness of using calibration data from the previous year (i.e. adopting field calibration from the current year did not significantly affect block-level estimates). Calibration factors were similar to that reported by Stein *et al.* [3] for Orchard 1, meaning factors did not change significantly over three consecutive seasons, which is encouraging for using the system as a 'black box'.

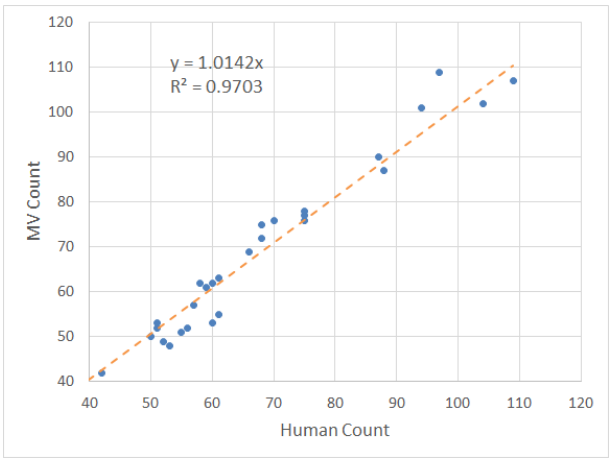
Platform B. For the YOLO model, the best F1 score (0.951) was obtained for the training set with a Non-maximum

suppression (NMS) threshold of 0.4 and class confidence cut-off threshold of 0.257 (data not shown). The correlation between human count of fruit per image and the machine vision count was strong (e.g. R^2=0.97, slope=1.01 for the Orchard 1 calibration tree set, Figure 4b), indicative of high accuracy in the machine vision count. This result was comparable to that obtained with the VGG F-RCNN for these night images, but the YOLO model operated at an order of magnitude higher speed (data not shown).

However, the correlation of dual view machine vision estimate number of fruits to harvest number was characterized by a R^2 between 0.64 and 0.85 and a slope between 0.37 and 0.69 for the orchards considered (data not shown). Thus, while the technique was successful in detecting exposed and partially occluded fruit, a significant proportion of fruit on the tree remain unseen in the dual view technique. Thus a correction factor for occluded fruit was employed, based on total fruit numbers of the calibration trees.



(a)



(b)

Figure 4. Regression of (a) fruit counts estimates per tree from platform A machine vision system against manual field counts in Orchard 1; (b) platform B YOLO v3 estimate against human count of fruit in tree images.

C. Orchard yield estimate validation

The orchard fruit load estimates were compared to harvest packline data and to an estimate based on the count of a subsample of (calibration) trees (Table 1). The subsample method involved a high human labour input, and the estimates varied between 97 and 150% of packhouse totals.

The weighted average percentage error (weighted by packhouse fruit numbers) was 6.0, 8.8 and 9.9% of packhouse counts for the day imaging system, night imaging machine vision system and the satellite method, respectively, averaged across all orchards assessed (data of Table 1).

All imaging methods achieved a good result on Orchard 1, with poorer results obtained on other orchards. The size and open canopy structure of Orchard 1 particularly suited in-field imaging, in comparison to other orchards with either smaller canopies (more chance for double counting of fruit) or larger canopies (more occlusion of fruit).

The satellite imagery procedure suffered a high uncertainty in the calibration of tree load to VI (Fig. 3), but the accuracy of the total orchard predictions was consistently high (Table 1). Note that estimates were based on imagery collected five months before harvest. To guide farm management decisions, calibration could be based on human count of fruit on tree at a time well before harvest.

Dual-view platform B delivered orchard predictions that over-estimated actual harvest, with underestimation due to occluded fruit exceeded by overestimation due to double counting from the two sides of the tree. The multi-view machine vision method decreased count errors associated with occluded fruit and double counting of fruit from imaging each tree from both inter-row sides, thus reducing the reliance on a manually estimated correction factor per orchard. The model was trained using image snips from one orchard only, and that from the previous season, with the predictions thus demonstrating model robustness across season, orchard and location.

With an industry move to higher density, smaller tree plantings, this method offers potential for use without need to estimate a correction factor per orchard. Additional to orchard fruit load estimation, these methods can monitor the yield of individual trees over seasons, enabling the identification of elite trees.

TABLE 1. HARVEST PACKHOUSE COUNT OF FRUIT AND ESTIMATION FROM SEVERAL METHODS INVOLVING COUNT OF A SAMPLE OF TREES OR SATELLITE IMAGERY VI AND MULTI-VIEW/DUAL-VIEW MACHINE VISION OF THE ENTIRE ORCHARD. VALUES IN PERCENTAGE ARE ESTIMATED FRUIT NUMBER AS A % OF HARVEST NUMBER. THE VALUE IN BRACKETS IN THE TREE # COLUMN IS THE NUMBER OF TREES HARVESTS FOR THE SUB-SAMPLE ESTIMATE. THE SUBSAMPLE METHOD REFERS TO USE OF THE AVERAGE OF FRUIT NUMBERS FOR THE CALIBRATION TREES, MULTIPLIED BY TREE NUMBER PER ORCHARD.

Orchard	Tree #	Packhouse count	Sum-sample	WVIII	Multi-view (Platform A)	Dual-view (Platform A)	Dual-view (Platform B)
1 Calypso	494 (44)	97,382	94,700 (97%)	94,293 (97%)	94,115 (97%)	92119 (95%)	100,217 (103%)
2 Calypso	386 (54)	67,110	75,810 (113%)	59,552 (89%)	61,279 (91%)	54912 (82%)	74,435 (111%)
3 Calypso	980 (39)	175,431	171,794 (98%)	201,036 (115%)	163,709 (93%)	149488 (85%)	-
4 R2E2	78 (18)	2,110	3,159 (150%)	-	2,693 (128%)	3,860 (183%)	2,701 (128%)
4 R2E2	224 (18)	21,180	24,086 (114%)	20,891 (99%)	21033 (100%)	27003 (128%)	-
5 HG A	616 (18)	36,490	44,557 (122%)	43,814 (107%)	-	-	47,248 (130%)
6 HG B	994 (23)	143,544	153,162 (107%)	-	-	-	152,756 (106%)

D. *In field fruit sizing*

Size of un-occluded fruit was estimated from pixel dimensions and estimated camera to fruit distance using Platform B (imaged one day before harvest). The RMSE on estimated length and width was 5 mm relative to caliper measurements (data not shown). The length and width dimensions were converted to weight estimates using an allometric relationship ($W = 0.49*L*W^2$), and to tray size based on use of a 7.5 kg tray. The resulting population distribution, derived from all non-occluded fruit in images of all trees in the orchard, was comparable to that of the packline data (Figure 5). Practical use would involve measurement of fruit size some weeks before harvest, with an assumed rate of increase applied to estimate size distribution at harvest.

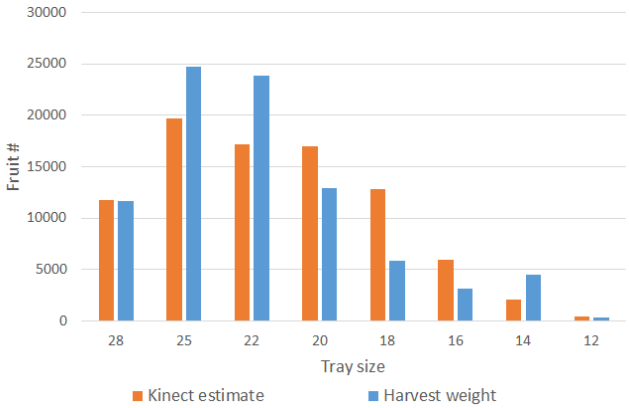


Figure 5. Packhouse (blue) and in-field image (orange) based estimates of the population distribution based on tray size of fruit, for Orchard 1.

IV. CONCLUSION

The three studied approaches were all able to provide sufficiently accurate estimates to support farm management and decision making. Satellite imagery allows for area wide assessments. Estimation was based on imagery collected five months before harvest in this study, although orchard level calibration is required using fruit counts of calibration trees at a time closer to harvest. Reliability of in-field machine vision estimates of fruit load was improved with use of the multi-view technique, with need for a correction factor for occluded fruit likely to be removed for orchards with narrow canopies. Note that for in-field machine vision counts, the whole orchard need not be assessed, but rather a sample size consistent with the required precision. Estimates of fruit load per tree accumulated across years may be useful for selective breeding and to inform management interventions. Fruit size estimation complements fruit load estimation to inform harvest management decisions. These developments in fruit detection, localization and sizing also set the scene for a robotic harvest operation.

ACKNOWLEDGMENT

Support of Simpsons and Philpots farms, Childers, Queensland, Groves Grown farm, Bungundara, is appreciated. This work was supported by the Australian Centre for Field Robotics (ACFR) at The University of Sydney, a CQUniversity RUN scholarship to AK (with co-supervision of C. McCarthy of Uni Southern Qld) and CQUni fellowship to ZW, and the Precision Agriculture Group at University of New England. Funding support (grants ST15002, ST15005 and ST15006) from Hort. Innovation Australia and the Australian Government Department of Agriculture and Water Resources as part of its Rural R&D Profit program is acknowledged.

REFERENCES

[1] A. Robson, M. Rahman, and J. Muir, "Using worldview satellite imagery to map yield in avocado (Persea americana): a case study

in Bundaberg, Australia," *Remote Sensing*, vol. 9, no. 12, p. 1223, 2017.

[2] A. Payne, K. Walsh, P. Subedi, and D. Jarvis, "Estimating mango crop yield using image analysis using fruit at 'stone hardening' stage and night time imaging," *Computers and Electronics in Agriculture*, vol. 100, pp. 160-167, 1// 2014.

[3] M. Stein, S. Bargoti, and J. Underwood, "Image based mango fruit detection, localisation and yield estimation using multiple view geometry," *Sensors*, vol. 16, no. 11, p. 1915, 2016.

[4] Z. Wang, K. B. Walsh, and B. Verma, "On-tree mango fruit size estimation using RGB-D images," *Sensors*, vol. 17, no. 12, p. 2738, 2017.

[5] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3626-3633.

[6] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23-27, 1975.

[7] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[8] W. Spreer and J. Müller, "Estimating the mass of mango fruit (Mangifera indica, cv. Chok Anan) from its geometric dimensions by optical measurement," *Computers and Electronics in Agriculture*, vol. 75, no. 1, pp. 125-131, 1// 2011.

[9] N. T. Anderson, P. P. Subedi, and K. B. Walsh, "Manipulation of mango fruit dry matter content to improve eating quality," *Scientia Horticulturae*, vol. 226, pp. 316-321, 2017.

[10] A. Robson, J. Petty, D. Joyce, J. Marques, and P. Hofman, "High resolution remote sensing, GIS and Google Earth for avocado fruit quality mapping and tree number auditing," in *Proceedings of the 29th International Horticultural Congress, Brisbane Convention and Exhibition Centre, Brisbane, Australia*, 2014, pp. 17-22.