Computers and Electronics in Agriculture 214 (2023) 108302

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples

Juan C. Miranda [a,*], Jaume Arnó [a], Jordi Gené-Mola [a,b], Jaume Lordan [c], Luis Asín [c], Eduard Gregorio [a]

[a] Research Group in AgroICT & Precision Agriculture, Department of Agricultural and Forest Sciences and Engineering, Universitat de Lleida – Agrotecnio CERCA Centre, 25198 Lleida, Catalonia, Spain
[b] Efficient Use of Water in Agriculture Program, Institute of AgriFood, Research and Technology (IRTA), Parc Científic i Tecnològic Agroalimentari de Gardeny (PCiTAL), Fruitcentre, 25003 Lleida, Spain
[c] Fruit Production Program, Institute of AgriFood, Research and Technology (IRTA), Parc Científic i Tecnològic Agroalimentari de Gardeny (PCiTAL), Fruitcentre, 25003 Lleida, Spain

ARTICLE INFO

Keywords:
Azure Kinect
Fruit sizing
Allometric weight models
Apple tree
Digital fruit growing

ABSTRACT

Data acquired using an RGB-D Azure Kinect DK camera were used to assess different automatic algorithms to estimate the size, and predict the weight of non-occluded and occluded apples. The programming of the algorithms included: (i) the extraction of images of regions of interest (ROI) using manual delimitation of bounding boxes or binary masks; (ii) estimating the lengths of the major and minor geometric axes for the purpose of apple sizing; and (iii) predicting the final weight by allometric modelling. In addition to the use of bounding boxes, the algorithms also allowed other post-mask settings (circles, ellipses and rotated rectangles) to be implemented, and different depth options (distance between the RGB-D camera and the fruits detected) for subsequent sizing through the application of the thin lens theory. Both linear and nonlinear allometric models demonstrated the ability to predict apple weight with a high degree of accuracy (R^2 greater than 0.942 and RMSE < 16 g). With respect to non-occluded apples, the best weight predictions were achieved using a linear allometric model including both the major and minor axes of the apples as predictors. The mean absolute percentage error (MAPE) ranged from 5.1% to 5.7% with respective RMSE of 11.09 g and 13.02 g, depending to whether circles, ellipses, or bounding boxes were used to adjust fruit shape. The results were therefore promising and open up the possibility of implementing reliable in-field apple measurements in real time. Importantly, final weight prediction error and intermediate size estimation errors (from sizing algorithms) interact but in a way that is not easily quantifiable when weight allometric models with implicit prediction error are used. In addition, allometric models should be reviewed when applied to other apple cultivars, fruit development stages or even for different fruit growth conditions depending on canopy management.

1. Introduction

Fruit size and weight are important quality parameters which strongly affect the final price of fruit. Monitoring these parameters throughout the season provides invaluable information (such as. growth curves) to support decision making in fruit crop management (Alibabaei et al., 2022). Knowledge of fruit size and weight is also key to making accurate yield predictions which allow fruit growers to plan the resources required (labour, transport, cold rooms) during harvesting, design marketing strategies and, ultimately, contribute to optimizing orchard profitability (Anderson et al., 2021; He et al., 2022).

At present, estimates of fruit size tend to be based on manual measurements, involving the use of Vernier callipers or sizing rings on a sample of trees. This is a labour-intensive and time-consuming approach whose practical application is both difficult and susceptible to errors. To overcome these limitations, several automatic methods for in-field fruit size estimation have been proposed, which can be classified depending on the type of data used (2D images and 3D point clouds). Information about fruit size can be extracted from 2D images, either by using calibration targets of a known size *in situ* (Wang et al., 2020) or by measuring the distance between the camera and the fruits in a given image (Gongal et al., 2018). More recently, it has become possible to

* Corresponding author.
E-mail address: juancarlos.miranda@udl.cat (J.C. Miranda).

https://doi.org/10.1016/j.compag.2023.108302
Received 3 May 2023; Received in revised form 19 September 2023; Accepted 30 September 2023
Available online 7 October 2023

generate point cloud reconstructions of fruits and to measure their size by applying 3D sensing techniques, such as light detection and ranging (LiDAR), photogrammetry techniques, or RGB-D cameras (Hacking et al., 2019; Tsoulias et al., 2020; Gené-Mola et al., 2021).

Of these 3D techniques, RGB-D cameras stand out for their transfer potential to the sector, due to their low cost and the ability to simultaneously provide colour, depth and infrared images at high acquisition rates (Fu et al., 2020; Gregorio and Llorens, 2021). One limitation is that they tend to provide poorer results under direct sunlight (Rosell-Polo et al., 2015; Gené-Mola et al., 2020a). RGB-D cameras have been used for in-field fruit sizing in crops including mango (Neupane et al., 2022; Wang et al., 2017), grape (Kurtser et al., 2020), apple (Mengoli et al., 2022) and pomegranate (Yu et al., 2022).

Over the years, different models have been proposed for assessing fruit weight, based on predicting fruit growth patterns (in weight) as a function of days after bloom (Mitchell, 1986; Lakso et al., 1995). The performance of these models has, however, often been affected by variability in meteorological conditions and management strategies. Another conventional approach is based on allometric relationships between fruit weight and geometric features. Amongst others, these features include: apple (Welte, 1990; Stajnko et al., 2013; Marini et al., 2019) and pear (Mitchell, 1986) diameter, the minor diameter in apple (Tabatabaeefar and Rajabipour, 2005) and pomegranate (Khoshnam et al., 2007), the perimeter in peach (Dalmases et al., 1998), and the length, maximum width and maximum thickness in mango (Spreer and Müller, 2011). An excellent summary of allometric relationships between fruit weight and linear dimensions can be found in Neupane et al. (2023).

In the current work, an automatic methodology is proposed for the in-field prediction of apple fruit size and weight. Colour and depth images, provided by an RGB-D camera, were used to study a set of apples that were manually labelled, simulating a perfect detector. The proposed methodology has a modular structure and allows the combined use of: different fruit-shape fittings; different methods for estimating fruit to camera distance; and different allometric weight models. As well as counting fruits, there is an increasing demand for ways of providing reliable estimates of yield per plot or per hectare. Hence the need for, and purpose of, this research, whose aim is to evaluate different sizing algorithms and allometric models and to provide the best possible way to complement currently available fruit detectors. The difficulty lies in combining the two tasks of lineal dimensions' estimation and weight prediction from lineal dimensions within a single, reliable, sequential automatic procedure.

2. Materials and methods

Fig. 1 provides a schematic view of the information flow between the three blocks on which the present research is based: i) data acquisition, ii) dataset creation, and iii) fruit size and fruit weight prediction. Data acquisition was carried out in an apple orchard, after previously selecting 12 trees in a given row. Before harvesting, video records were taken on three of the twelve trees (specifically, those numbered 1 to 3) using an RGB-D camera from a fixed platform ('fruit trees data acquisition', Fig. 1). Then, at harvest, the fruits from the 12 selected trees were collected and characterized in the laboratory, with their size and weight being individually determined ('fruit characterization in laboratory', Fig. 1). The second block is related to the creation of the data set. Videos recorded in trees 1 to 3 mentioned above were processed to obtain images and create a dataset (n = 26) with manually labeled apples ('dataset construction and manual annotation', Fig. 1). In parallel, several allometric models for apple weight prediction were obtained based on the rest of the laboratory data (i.e., using information on fruits from trees 4 to 12) ('allometric weight modeling', Fig. 1). The third block involved applying the sizing algorithms and the proposed allometric models. This was first done separately and then combined sequentially ('prediction algorithms', Fig. 1). In a final step, the performance of the proposed algorithms was evaluated by contrasting several different statistical metrics ('evaluation and testing', Fig. 1).

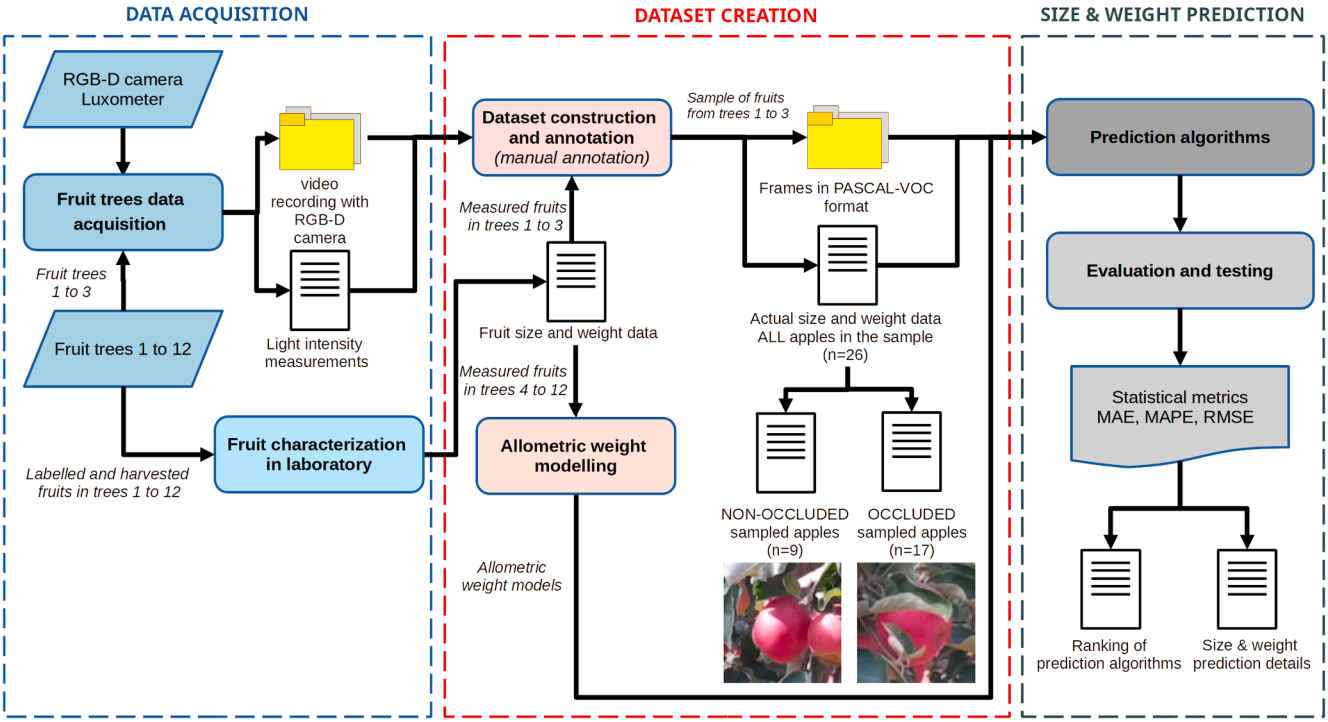


Fig. 1. Sequential methodology for the prediction of apple fruit size and weight using data collected with an RGB-D camera under field conditions. The three blocks (delimited by dotted lines) make up the global procedure, with each one including the different steps performed (rounded boxes) and highlighting the input and output data required and provided in each case.

2.1. Fruit-tree data acquisition

Field tests were carried out at an experimental apple orchard (cultivar Story® Inored[cov]) located in the municipality of Mollerussa, Catalonia, Spain (latitude: 41.617465 N; longitude: 0.870730; 246.3 m a.s.l. ETRS89) and owned by the Institut de Recerca i Tecnologia Agroalimentàries (IRTA). The trees in this orchard were trained as a fruiting wall, with a planting spacing of 3.6 × 1 m, and a maximum canopy height of 3.5 m. A set of 12 consecutive trees was selected for the study (Fig. 2a,b). RGB-D data acquisition was performed on three of them (trees 1 to 3), while the fruits from the remaining trees (4 to 12) were used to create allometric weight models.

The RGB-D camera used in these tests was the Azure Kinect DK (Microsoft Corporation, Redmond, WA, USA). This device combines a 1-megapixel time-of-flight (ToF) camera, a CMOS rolling shutter sensor, an inertial measurement unit (IMU) and a microphone array. In our experiment, the Azure Kinect camera was configured to save RGB, IR and depth data, while the IMU sensor and the microphone were disabled. The selected depth camera mode was narrow field-of-view (NFOV) unbinned, with the specifications shown in Table 1 (Microsoft, 2022).

The Azure Kinect camera was positioned so that it faced westward, with view of the canopy of a north–south oriented tree row. It was mounted on a tripod, at a height of 1.38 m, and it was positioned 1.50 m from the tree row axis (Fig. 2c,d,e). A Modern 15 A10RBS-484XES laptop (MSI, New Taipei, Taiwan), running Windows 10, was used as the host for the camera operation and data storage. As shown in Fig. 2f,

Table 1
Azure Kinect camera specifications provided by the manufacturer.

RGB frame resolution	1920 × 1080 pixels
RGB frame rate	30 fps
RGB field of view	90° × 59°
Depth frame resolution	640 × 576 pixels
Depth frame rate	30 fps
Depth field of view	75° × 65°
Depth range	0.5–3.86 m

two calibration foam spheres (60 and 120 mm diameter) were hung from a steel wire between trees 2 and 3, as was a DVM1300 digital light meter (Velleman, Gavere, Belgium), which was used to measure the illuminance throughout the experiment.

Data were acquired from 11:40 to 19:24 (UTC + 2) on September 27, 2021, when the fruit trees were at an advanced ripening growth stage BBCH 85 (Meier, 2018). The apples had starch indexes of 8–9 (1–10 scale), soluble solids contents of 8.2° Brix, and firmness values of 14.6 kg/cm^2. A total of 25 video recordings (one every 15–20 min), each with a duration of 4 s, were recorded, using AK_ACQS software (Miranda et al., 2022). The illuminance was registered for each video capture, with prevailing sunny conditions in the morning and slightly cloudy conditions throughout the afternoon.

On September 29, 2021, after data acquisition, the fruits from the 12 selected trees were labelled in the field (Fig. 3a), using adhesive paper stickers (Fig. 3b). Several videos, identification notes and photos were also taken as ancillary data, in order to keep evidence of fruit positions

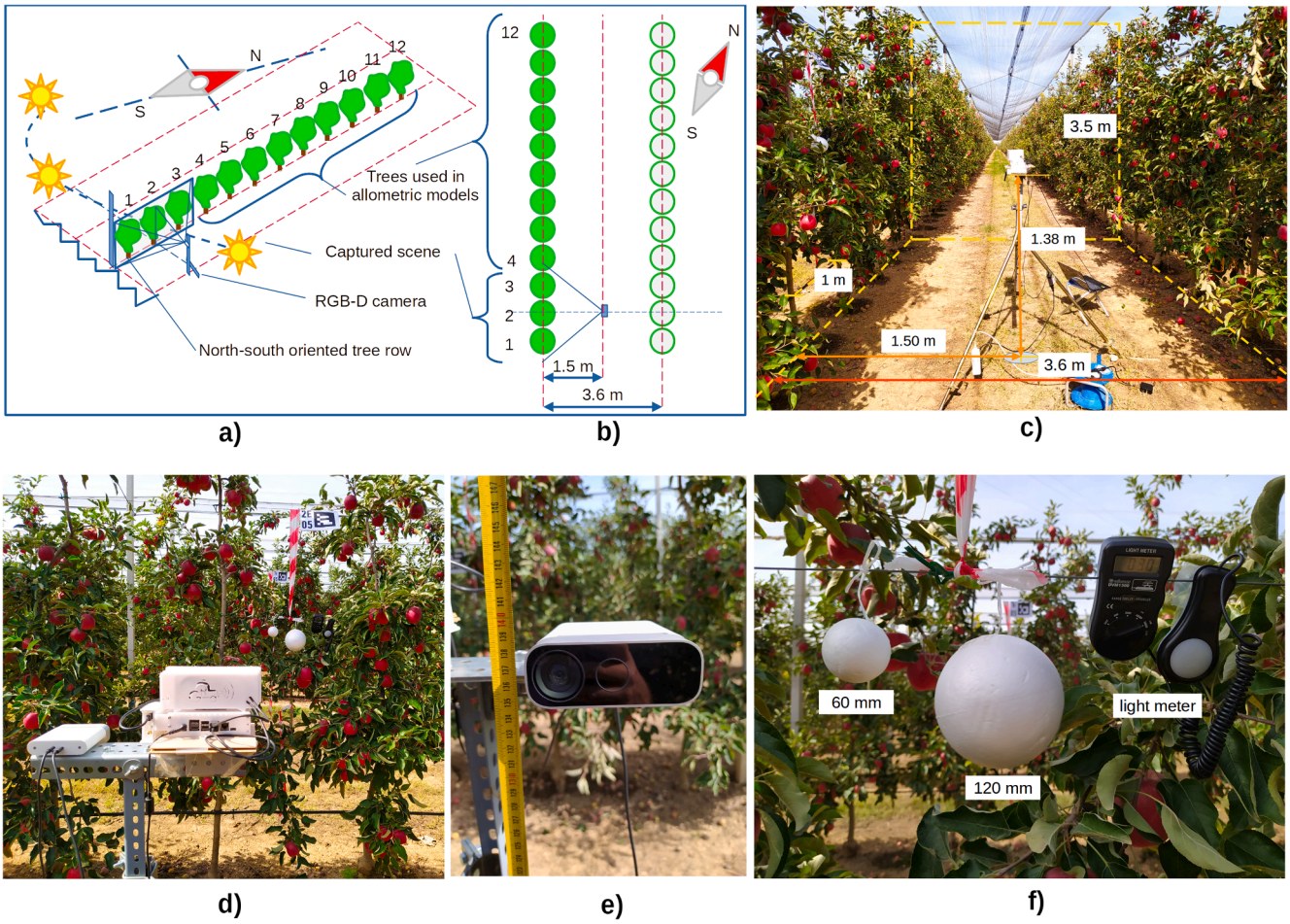


Fig. 2. Field experimental set-up. a) A perspective representation of the scene, showing the relative position of the sun throughout the experiment. b) Plan view of the layout. c) View of the tree alley, showing the planting pattern and camera position. d) View of the captured scene (trees 1 to 3). e) Azure Kinect camera. f) Calibration foam spheres and digital light meter placed on the trees.

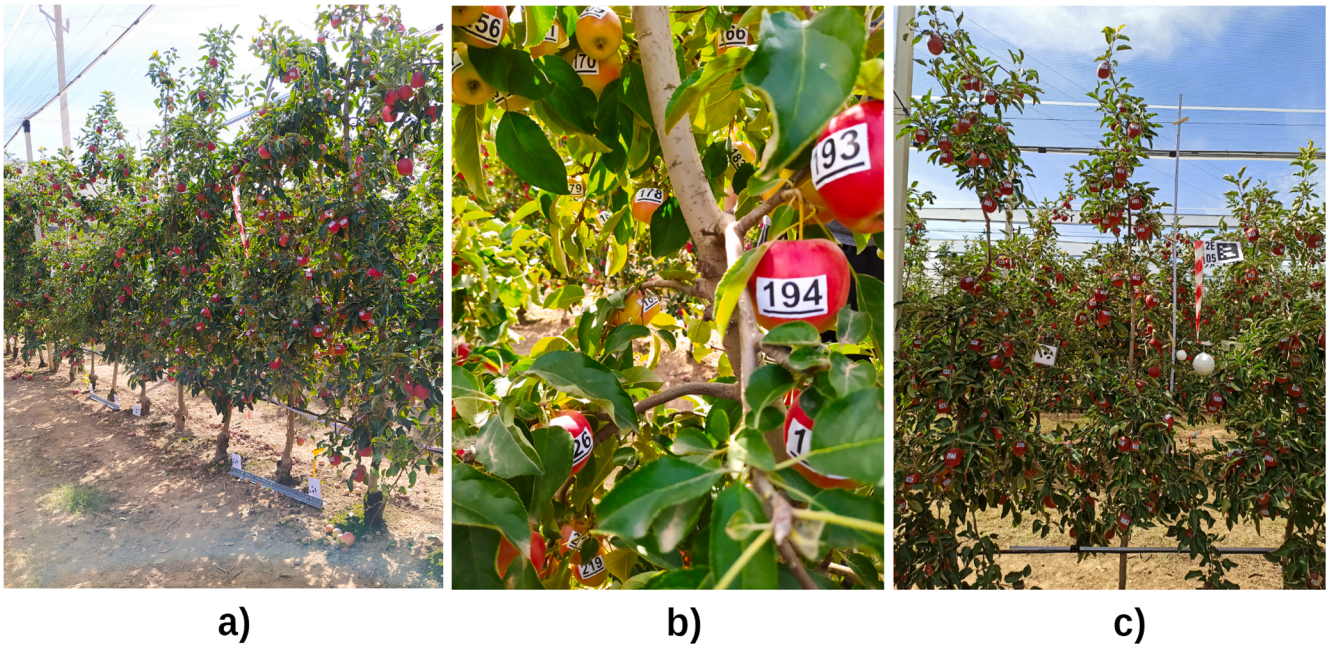


Fig. 3. Apple labelling. a) Overall view of the selected trees with labelled fruits. b) Detail view of the apples with their adhesive paper stickers. c) Front view of a single tree, used to identify the position of each labelled apple.

within each tree (Fig. 3.c). The fruits were hand-picked on September 30 and October 1, 2021 and placed in different collapsible plastic storage boxes (one for each tree).

2.2. Fruit characterization in the laboratory

A total of 1321 apples were harvested and stored in a cold room at 4 °C to conserve their organoleptic characteristics. In the following days, the boxes of fruit were moved out of the cold storage room and transported to the laboratory for fruit characterization (Fig. 4a).

The dimensions of each apple were measured, using Vernier callipers, placing the stem upwards and recording its calibre (or width) (C) and height (H) (Fig. 4b). A CB 501 weighing machine (Adam Equipment, Oxford, CT, USA) was used to measure fruit weight (W) (Fig. 4c). Then, the size measurements of each apple (C and H) were then compared to each other to create two new data fields, in which the largest measurement was recorded as D_1 and the smallest as D_2. The resulting fruit size and weight data (D_1, D_2, W) were used to create a database organized using tree and apple identifiers.

2.3. Allometric weight modelling

Linear and nonlinear models were fitted, taking the geometric measurements of the apples (D_1 and D_2) as predictors and the weight of the apples (W) as the response variable. Mathematically speaking, one very general form for the model would be: $W = f(D_1, D_2) + \varepsilon$, where f is an unknown function and ε is the error term (or residual). Linear models were considered amongst the different possible functions, primarily because of the empirical nature of the research.

The first model considered using only the largest measured dimension of the fruit (axis D_1) as a predictor. A simple linear regression was therefore tested as: $W = \beta_0 + \beta_1 D_1 + \varepsilon$, with β_0 and β_1 as the unknown

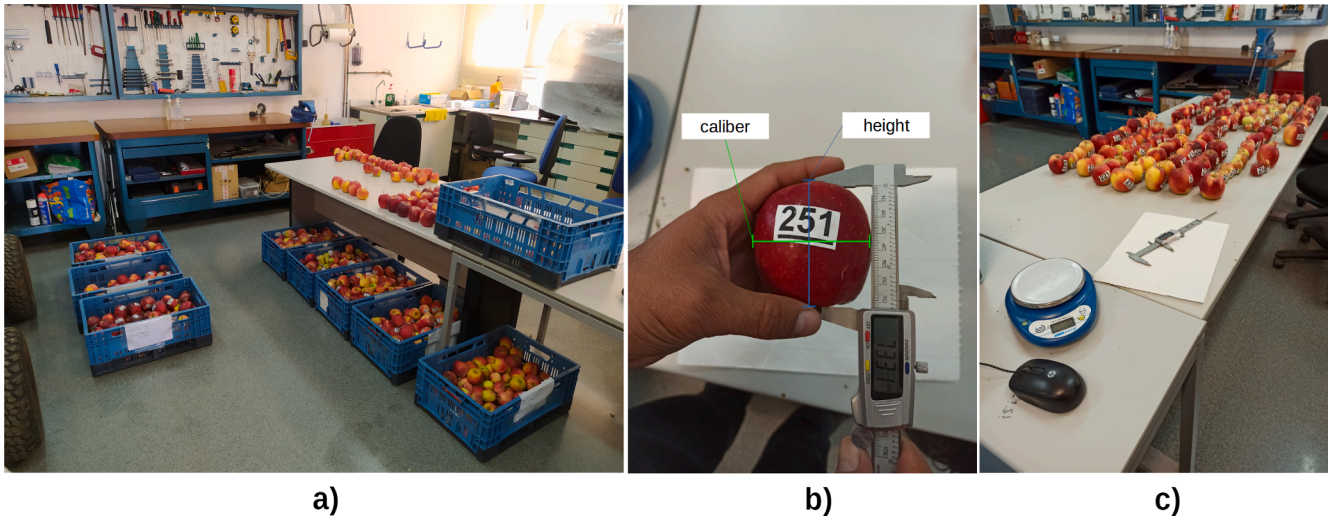


Fig. 4. Characterization of fruits in the laboratory. a) Fruits in storage boxes, identified by tree. b) Apple size measurement: calibre (horizontal axis) and height (vertical axis). c) Measured and weighed fruits used to create an organised database.

parameters of the model. The addition of polynomial terms in this same single-predictor case (predictor D_1) allowed us to test a second model: $W = \beta_0 + \beta_1 D_1 + \cdots + \beta_d D_1^d + \varepsilon$ for a more flexible relationship. The exponent d was chosen until we obtained a term $d + 1$ that was not statistically significant.

The third model tested was also linear, but used the two geometric measures (D_1 and D_2) as predictors: $W = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \varepsilon$. It was expected that this model would improve the predictions with respect to the first one. However, problems of collinearity, leading to imprecise estimates of β, were also expected using this model given the almost certain relationship between the two geometric fruit measurements (D_1 and D_2) that were used as predictors. To check whether the two predictors should be used together, the variance inflation factor (VIF) $\left(1 - R_j^2\right)^{-1}$ was calculated (in which R_j^2 is the coefficient of determination of the linear regression between D_1 and D_2). Statistically speaking, a high value of this factor would make it advisable to remove one of the predictors from the model (Faraway, 2016), with it being more reasonable to predict the weight of the apples from a single geometric measurement of the fruit.

The fourth and fifth models, although linear in terms of their parameters, included a specific combination of the D_1 and D_2 measurements as the sole predictor, considering that apples, as 3D objects, could be roughly adjusted to the volume of a sphere or an ellipsoid. Seeking to combine the two diameters in order to achieve a magnitude that could be measured as a unit of volume (the magnitude of cubic length), the simplest possible models were: $W = \beta_0 + \beta_1 \left(D_1^2 D_2\right) + \varepsilon$, and $W = \beta_0 + \beta_1 \left(D_1 D_2^2\right) + \varepsilon$. Finally, the respective nonlinear models: $W = \beta_0 \times D_1^{\beta_1} + \varepsilon$ and $W = \beta_0 \times D_1^{\beta_1} \times D_2^{\beta_2} + \varepsilon$, were taken as the sixth and seventh options to be assessed. By applying the appropriate transformation, both nonlinear models were linearized in order to better estimate their β parameters.

A least squares estimation was used to estimate the β parameters for the seven allometric models cited above, while the goodness-of-fit was assessed using the coefficient of determination R^2 in each case. To avoid relying solely on R^2 as a measure of fit, the root mean square error (*RMSE*) was also used. To be more specific, models were obtained from a training dataset (568 apples) in order to subsequently check the *RMSE* obtained when the models were applied to a test dataset (489 apples) which had been constructed using the rest of the 1057 fruits which had been collected from trees 4 to 12 (see Figs. 1 and 2). In addition to all of the above, we also performed diagnostics on the assumption of homoscedasticity and normality of the residuals for each of the proposed models. The associated allometric modelling was carried out using RStudio version 1.4.1717 software.

2.4. Dataset construction and annotation (manual annotation)

As previously stated, the goal of this work was to propose and assess different fruit size and weight prediction algorithms based on RGB-D data. As a result, and with the aim of avoiding potential errors that could arise from the object detection process, the fruits were manually labelled to emulate high accuracy detection. The creation of the labelled dataset was divided into four steps: 1) frame extraction; 2) object annotation and file conversion; 3) binary mask creation; and 4) the checking of fruit label location.

Firstly, five video recordings made in the morning (11:40, 11:59, 12:18, 12:35, 12:53 UTC + 2) were selected together with one that was made in the late afternoon (19:24 UTC + 2). The morning videos corresponded to the best lighting conditions of the scene (trees 1 to 3) considering that it was east facing. The afternoon video was selected in order to assess how backlighting conditions (sunset) affected RGB-D measurements. One frame per video (taken 1 s after the video starting) was extracted using the AK_FRAEX software (Miranda et al., 2022). From it, RGB, IR and depth registered images were obtained.

Secondly, object annotation was performed on the RGB images. The positions of 26 apples and 2 calibration foam spheres were labelled on each of the images using the Pychet Labeller tool (Bargoti, 2016) configured for bounding box markings (Fig. 5). Apples within the field of view of the depth camera were considered to construct a dataset including non-occluded and occluded fruits, depending on whether they were completely visible (or almost) or only partially visible, respectively. Decision on which occluded fruits could be labeled was made by two technical specialists in this area, that is, based on their experience and without setting any maximum level of occlusion. Then, file conversion from plain text to PASCAL-VOC format (Everingham et al., 2010) was done to create correspondence files between each frame (image) and its annotations.

Thirdly, a binary mask was generated from each RGB image, using the Matlab® Image Segmenter tool (version R2021a, MathWorks Inc., Nastick, MA, USA) to delimitate the object regions in pixels.

Then, fruit label location in images was checked by comparison with a photogrammetry-based 3D reconstruction. To do this, a 3D point cloud of the scene (trees 1 to 3) was created using a Canon EOS 60D DSLR Camera (Canon Inc., Tokyo, Japan), following the methodology proposed by Gené-Mola et al. (2020b). Ancillary video data for labelling verification was provided by a Redmi Note 8 T mobile phone (Xiaomi, Beijing, China). As a result of the previous steps, a hierarchical metadata folder containing RGB, IR and depth images, object annotations and binary masks was created for each frame extracted from the initial video set.

On the basis of the laboratory measurements (Section 2.2) and the fruits identified in the labelling process (Section 2.4), actual size and weight data for all the apples in the sample was created and saved in a general set (ALL). As shown in Fig. 1, apples within the dataset were grouped into two subsets, non-occluded apples (n = 9) and occluded apples (n = 17).

2.5. Prediction algorithms

Fig. 6 provides an overview of the algorithms for fruit size and weight prediction developed in this work. Image datasets in PASCAL-VOC format (Section 2.4), which include RGB images, depth images and binary masks, were used as input for the prediction algorithms (Fig. 6a). Two different approaches were then used to identify the regions of interest (ROI): i) bounding boxes (BBOX) (Fig. 6b.1); and ii) binary masks (MASK) (Fig. 6b.2). Both approaches included the following steps: 1) size estimation in pixels; 2) depth estimation; and 3) fruit size estimation. Finally, the allometric models inferred in Section 2.3 were applied to predict fruit weight (Fig. 6c).

The prediction algorithms were implemented using Python 3.8, Tkinter, and OpenCV for image processing, and also other open source libraries/packages, such as Numpy, Scikitlearn and Pandas. As result, a software package with graphic user interfaces was published as Python package containing the pipeline implemented (Miranda et al., 2023).

2.5.1. Size estimation in pixels

At this step, pixel lengths of the major (D_1) and minor (D_2) axes of each fruit were extracted from images. In the bounding box approach, the box sides were used to estimate the lengths of the fruit axes (Fig. 7). This is a pixel-sizing method that has the advantage of being directly applicable when used with the most common bounding box-based object detectors.

In contrast, in the binary mask approach, the images were first smoothed, by applying morphological erosion and dilation operators (5 iterations and a 3 × 3 kernel), and fruit region contours were then detected. Once the contour points had been identified, the following shape-fitting techniques were assessed to estimate fruit size (D_1, D_2) for both non-occluded and occluded fruits (Fig. 7).

Fig. 5. Fruits selected and labelled from trees 1 to 3, in an image taken at 19:24 UTC + 2. The hexagonal area indicates the field of view of the Azure Kinect depth camera for the NFOV operating mode, with RGB and image depth overlapping.

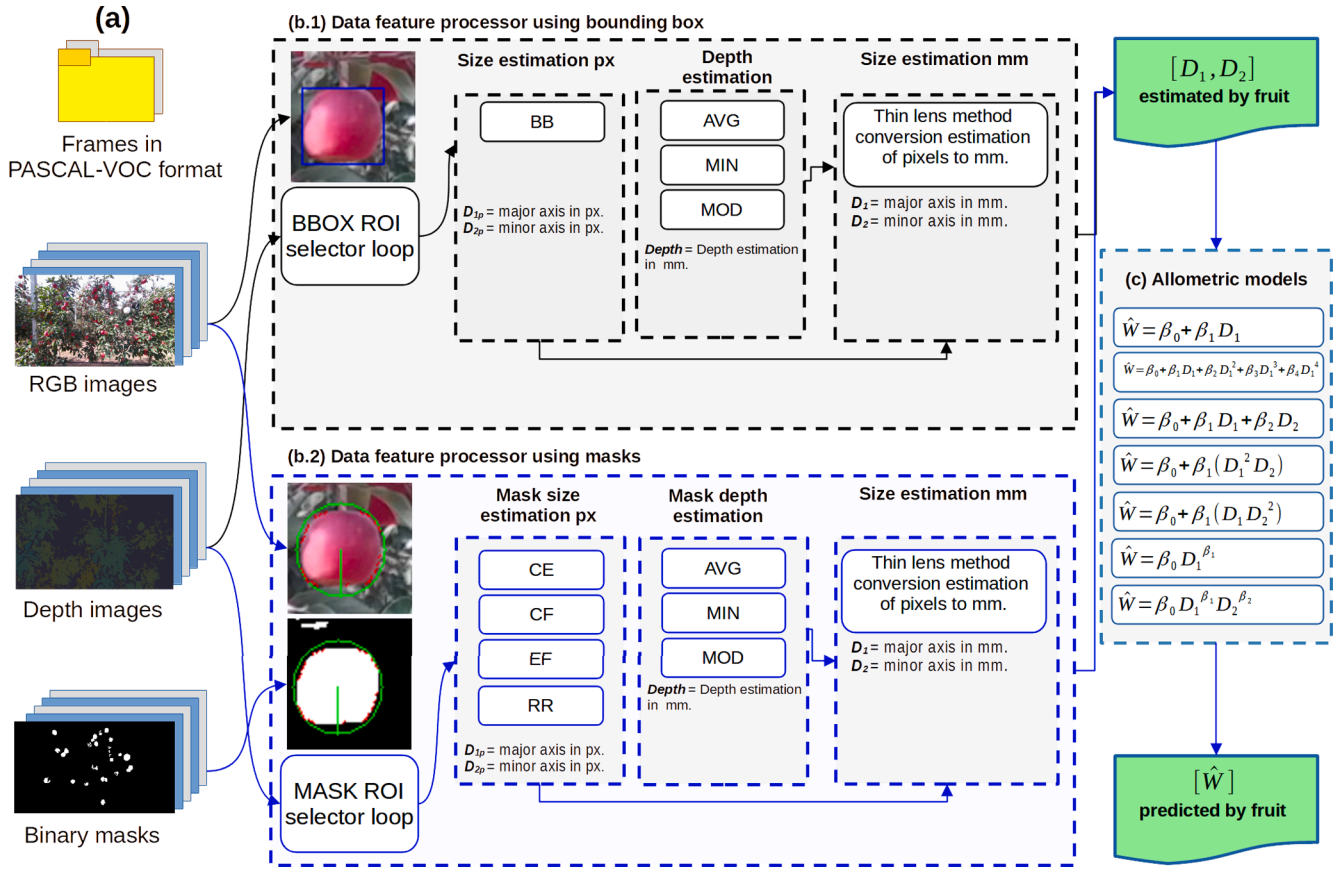
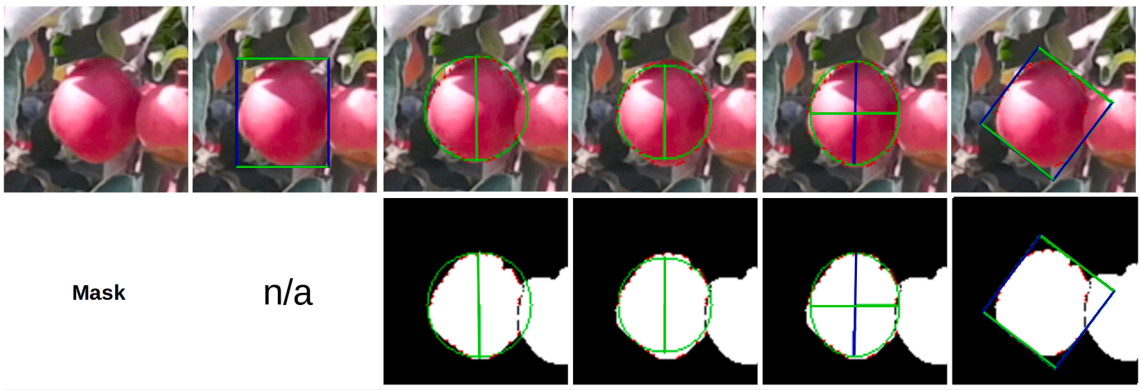
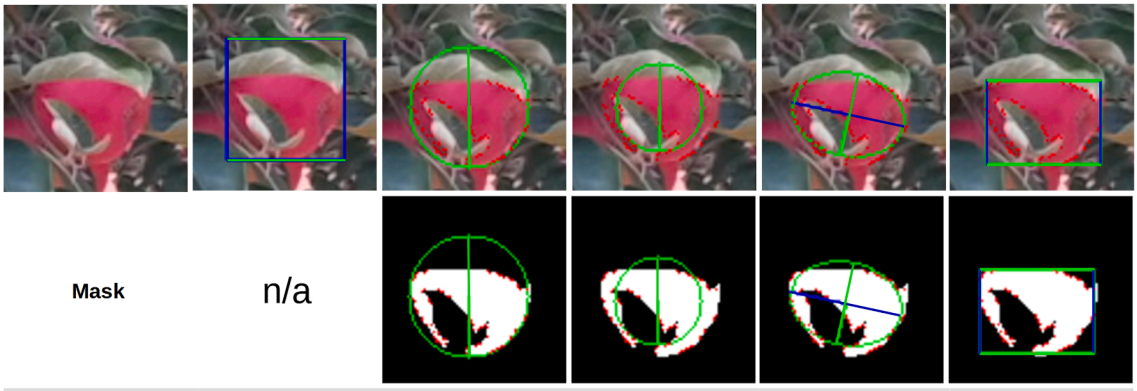


Fig. 6. Overview of the size and weight prediction algorithms applied. a) Input for prediction algorithms. Approaches for identifying regions of interest (ROI): b.1) bounding box (BBOX), b.2) binary masks (MASK). Size estimation: bounding box (BB), circle enclosing (CE), circle fitting (CF), ellipse fitting (EF), rotated rectangle (RR). Depth estimation: average (AVG), minimum (MIN), modal (MOD).

- Circle enclosing (*CE*): this computes a circumscribed circle that covers all the contour points within a minimum area.
- Circle fitting (*CF*): this fits a circle around the full list of contour points, using the least squares technique.
- Ellipse fitting (*EF*): this fits an ellipse to the contour points.
- Rotated rectangle (*RR*): this computes a rectangle with a minimal area which includes the contour points and considers the angle of its rotation.



Mask	n/a				
ROI extraction	BB	CE	CF	EF	RR
Size estimation [px]	D_{1p}=61 D_{2p}=52	D_{1p}=60 D_{2p}=60	D_{1p}=54 D_{2p}=54	D_{1p}=58 D_{2p}=51	D_{1p}=54 D_{2p}=52
Depth estimation [mm]	$Depth$=1209.73	$Depth$=1185.99	$Depth$=1185.99	$Depth$=1185.99	$Depth$=1185.99
Size estimation [mm]	D_1=70.96 D_2=60.49	D_1=68.42 D_2=68.42	D_1=61.58 D_2=61.58	D_1=66.14 D_2=58.16	D_1=61.58 D_2=59.30
Weight prediction model [g]	$\hat{W}$=148.61	$\hat{W}$=159.95	$\hat{W}$=127.79	$\hat{W}$=131.50	$\hat{W}$=122.72
a)	**b)**	**c)**	**d)**	**e)**	**f)**



Mask	n/a				
ROI extraction	BB	CE	CF	EF	RR
Size estimation [px]	D_{1p}=51 D_{2p}=51	D_{1p}=52 D_{2p}=52	D_{1p}=38 D_{2p}=38	D_{1p}=48 D_{2p}=35	D_{1p}=49 D_{2p}=35
Depth estimation [mm]	$Depth$=1127.76	$Depth$=1129.36	$Depth$=1129.36	$Depth$=1129.36	$Depth$=1129.36
Size estimation [mm]	D_1=55.30 D_2=55.30	D_1=56.46 D_2=56.46	D_1=41.26 D_2=41.26	D_1=52.12 D_2=38.0	D_1=53.21 D_2=38.0
Weight prediction model [g]	$\hat{W}$=98.28	$\hat{W}$=103.76	$\hat{W}$=32.30	$\hat{W}$=52.0	$\hat{W}$=54.69
a)	**b)**	**c)**	**d)**	**e)**	**f)**

Fig. 7. Size and weight estimates of: (TOP) non-occluded apple # 2167, C = 63.68mm, H = 66.07mm, W = 136.4g; (BOTTOM) occluded apple # 2171, C = 64.43mm, H = 54.57mm, W = 102.3g. The first row corresponds to the RGB images and the second to the binary mask. a) Original fruit images (taken at 12:35 UTC + 2). b) Bounding box (BB). c) Circle enclosing (CE). d) Circle fitting (CF). e) Ellipse fitting (EF). f) Rotated rectangle (RR). In BB and RR, the D_1 axis is in blue and the D_2 axis in green. For CE and CF, the radius is in green. In EF, the D_1 axis is in blue and the D_2, or minor axis, is in green. Depth estimation was obtained by averaging the depth values for the selected ROI; fruit weight was predicted using allometric model (3) in Table 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.5.2. Depth estimation

To estimate actual fruit sizes (in mm) from measurements in pixels (Section 2.5.1), it was necessary to know the distances (depth) between the Azure Kinect camera and the fruits on the trees. Depth images provided by the RGB-D camera were used to compute an estimated depth, in mm (*Depth*), for each fruit. In the bounding box approach, the depth was directly estimated based on the pixels in the depth image inside the box. In contrast, in the binary mask approach, only the pixels within the fruit region were considered. This region was identified by overlapping the depth image a with a binary mask (bitwise matrix multiplication). In both approaches, the depth estimation of each fruit was provided for three statistical metrics related to the selected ROI: the average (AVG), modal (MOD) and minimum (MIN) values. To avoid errors, pixels from the depth images with values of zero (resulting from reflections, multipath errors, or fading, etc.) were excluded from the calculation.

Fig. 7 shows the estimations of depth (average value) for apple # 2167, applying both the bounding box (BB) and binary mask (CE, CF, EF, RR) approaches. For the same apple, Fig. 8a represents the regions used for depth estimation (RGB image and binary mask), while Fig. 8b shows the 3D plots of the depth values in the selected regions for the BBOX and MASK approaches. When using a bounding box, outlier values (high red values in these figures) appeared due to the presence of leaves and other vegetative elements within the ROI. In the case of the mask, most of the outliers were removed, which should have yielded a more accurate depth estimation.

2.5.3. Estimations of fruit size and predictions of fruit weight

The thin lens theory was applied to convert fruit size from pixels to mm:

$$D_i = \frac{(D_{ip} \times Depth)}{f_p}; i = 1, 2 \quad (1)$$

where D_i is the major/minor axis of the fruit in mm, D_{ip} is the major/minor axis of the fruit in pixels, and *Depth* is the depth value of the fruit (distance from the camera) in mm. $f_p = 1040$ is the scaled camera focal length (in pixels), which was experimentally determined using calibration spheres.

The predicted fruit sizes D_1 and D_2 were used as input parameters for the allometric models (Section 3.2) to predict fruit weight ($\widehat{W}$) in grams per fruit. For example, Fig. 7 shows size and weight predictions (in mm) for the non-occluded apple # 2167 and for the occluded apple # 2171 when the BBOX approach and tested shape-fitting techniques (CE, CF, EF, RR) were used. In this example, the allometric model $\widehat{W} = \beta_0 + \beta_1 D_1 + \beta_2 D_2$ was applied to predict the weight.

2.6. Evaluation and testing

The reliability of the prediction algorithms was verified using different statistical metrics. In a first step (Section 3.3), estimates of the geometric measurements of apples (axes D_1 and D_2) were tested separately. Then, in a second step and using the same metrics (Section 3.4), the joint performance of the sizing algorithms and adjusted allometric models was tested to predict the weight of the apples; this was done using the previously estimated D_1 and D_2 axes as inputs.

The evaluation metrics are listed in.

Table 2, where $\widehat{y}_i$ represents the predicted values (axes D_1 or D_2 obtained from the size estimation algorithms or, where appropriate, the weight $\widehat{W}$), and y_i the corresponding real values obtained from laboratory measurements (Section 2.2).

In the case of estimations of size, up to 15 different predictive options were assessed and then ranked from lowest to highest MAPE. These 15 possible results were obtained from combining the different pixel size adjustment options (BBOX_BB, MASK_CE, MASK_CF, MASK_EF, MASK_RR) and the proposed options for estimating depth (AVG, MIN, MOD) (Fig. 6). At the same time, the final weight prediction algorithms were ranked from best to worst predictive performance based on their metrics (this was done after assessing each of the 15 options for estimating size in combination with each of the seven allometric models). In short, it was possible to quantify error propagation for the different prediction phases (fruit size and weight) and, more importantly, it was also possible to contrast the impact of size prediction errors on predictions of fruit weight according to the different allometric models.

3. Results

3.1. Sizing error and image acquisition timing

Six different moments of video data capture (image acquisition timing), recorded from 11:40 to 19:24 (UTC + 2) on September 27, 2021, were compared to contrast the sizing errors and changing lighting conditions registered during a typical day. The illuminance of the canopy as seen by the camera (from an east facing light meter) decreased through the monitored period (Fig. 9). The six moments of capture are marked, with the first five covering the period from 11:40 to 12:53 (UTC + 2), under good lighting conditions, and the other, registered at 19:24 (UTC + 2), relating to late afternoon and very different illuminance

Table 2
Prediction algorithm evaluation metrics.

Root Mean Square Error (*RMSE*)	$RMSE = \sqrt{\frac{\sum_1^n (\widehat{y}_i - y_i)^2}{n}}$	(2)
Mean Absolute Error (*MAE*)	$MAE = \frac{\sum_1^n \lvert\widehat{y}_i - y_i\rvert}{n}$	(3)
Mean Absolute Percentage Error (*MAPE*)	$MAPE = \frac{1}{n}\sum_1^n \left\lvert\frac{\widehat{y}_i - y_i}{y_i}\right\rvert$	(4)

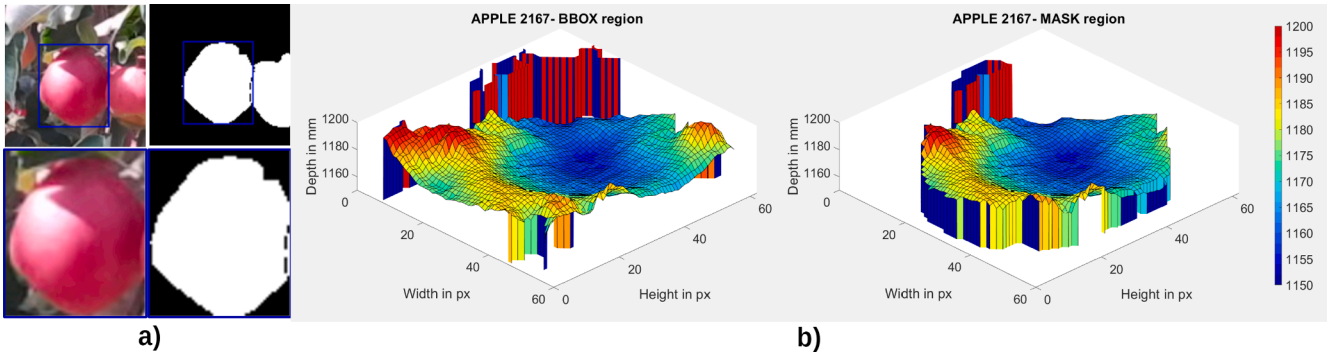


Fig. 8. Depth estimation of apple #2167 (taken at 12:35 UTC + 2), *BBOX average depth* = 1209.73 *mm*, *MASK average depth* = 1185.99 *mm*. a) RGB image and binary mask selected by bounding box rectangle. b) Depth values within the bounding box and mask region.

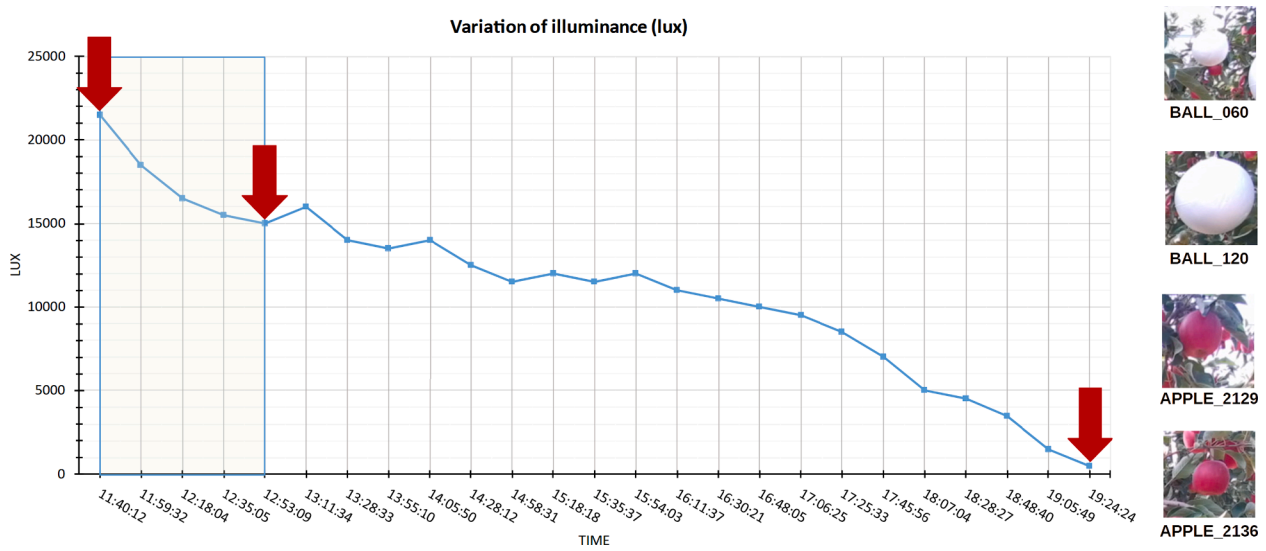


Fig. 9. Variation of illuminance (lux) at different times during the field data capture (September 27, 2021). The images of the spheres and apples correspond to the time slot 11:40:12 (UTC + 2). Red arrows represent the moments (delimiting the capture range or a specific moment) at which measurements were taken with the camera. The light meter was positioned to face eastwards, which explains why the values decrease and do not reach their maximum at noon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Section 2.4). For each of these times, sizing algorithms were applied to the captured images of two spheres (balls) of known sizes (60 and 120 mm in diameter) and also to two selected apples (#2129 and #2136) with known D_1 and D_2 axes.

Both ROI selectors (bounding box and mask) and the corresponding methods for estimating pixel size (BB, bounding box; CE, circle enclosing; CF, circle fitting; EF, ellipse fitting; RR, rotated rectangle) were used in this preliminary test. Depth estimations of objects using the average (AVG) method allowed estimates of fruit size (expressed in mm) to be obtained for a total of five possible outcomes for the spheres and apples. Figs. 10 and 11 (below) show the errors, comparing estimated and real measurements, for the different algorithm options and also for the six different moments of capture during the day.

At first sight, the variation in the degree of illuminance did not seem to significantly influence the estimations of diameter (size) made for the two spheres, when the sizing methods were applied individually (Fig. 10). Some methods clearly provided better estimates than others, with this being the case for one particular axis, regardless of lighting conditions. The methods that should perform better for fitting objects with different diameters D_1 and D_2 (ellipses and rotated rectangles) provide different errors for both axis. As expected, the methods based on circle fitting and circle enclosing provide similar results for both diameters of the calibration spheres.

Fig. 11 shows the results for the two apples chosen in the scene. As before, it was not possible to observe any clear pattern of errors associated with illuminance. In theory, therefore, almost any time window within daylight hours could have been chosen to use the RGB-D camera.

One particularly noteworthy result was that using a mask ROI selector in combination with the ellipse fitting (EF) sizing method provided the most reliable measurements for both the D_1 major axis and the D_2 minor axis. In contrast, the circle enclosing (CE) approach was found to be the least accurate sizing method (with very marked errors when estimating the minor axis D_2). This was because the CE method tends to fit circles that are outside the contour points, and which would correspond to the major axis (D_1); errors were therefore to be expected when estimating the length of the minor axis (D_2). In the rest of the sizing methods applied (BB, bounding box; CF, circle fitting; EF, ellipse fitting; RR, rotated rectangle), errors ranged between −6 mm and + 4 mm for both the D_1 and D_2 axes.

As relative errors may be considered acceptable (when considering the normal size of apples), there should have been no major problems involved in using RGB-D cameras while agricultural tasks were being performed. In the following sections, in-depth analyses were made of the allometric models and the data processing algorithms.

3.2. *Allometric models for predicting apple weight*

Table 3 shows the linear and nonlinear allometric models that were used to predict apple weight using the major and/or minor geometric axes of the fruit as predictors. The fit results were very good in all cases, with R^2 values ranging from 0.942 (simple linear model) to 0.993 (multiple nonlinear model). Any *a priori* choice between one model and another should therefore be based on some other criteria.

Models which produced low *RMSE* values in the training dataset and similar values to those obtained in the test dataset could be recommended. More specifically, the polynomial model had the advantage of using a single predictor (major axis D_1), resulting in the introduction of a single source of error into the model (this error related to the estimation of D_1). However, there was a possibility of amplifying the weight prediction error (model noise) as this predictor was used at different powers. Identical behaviour could have been expected, albeit to a lesser extent, in linear models based on the use of combined predictors, such as $D_1^2 D_2$, or - where appropriate - $D_1 D_2^2$. As for the nonlinear models, the use of both the D_1 and D_2 predictors provided the best weight predictions. However, once again, the estimation errors associated with these size predictors could have led to an amplified propagation of the error, given the potential use of exponents in the model (basically for the D_1 predictor).

Models that use the axes of the fruit as linear predictors, without the inclusion of exponents, should not, however, be ruled out. Error propagation in weight predictions could be lower in these models, even when the predictors are affected by higher regression coefficients. This was the case in models (1) and (3). However, we should also be cautious about using linear model (1) due to residual trend problems, and about opting for model (3) due to the existing correlation between predictors (VIF factor of 28.46) which, being greater than 10, indicates a high correlation and constitutes a cause for concern.

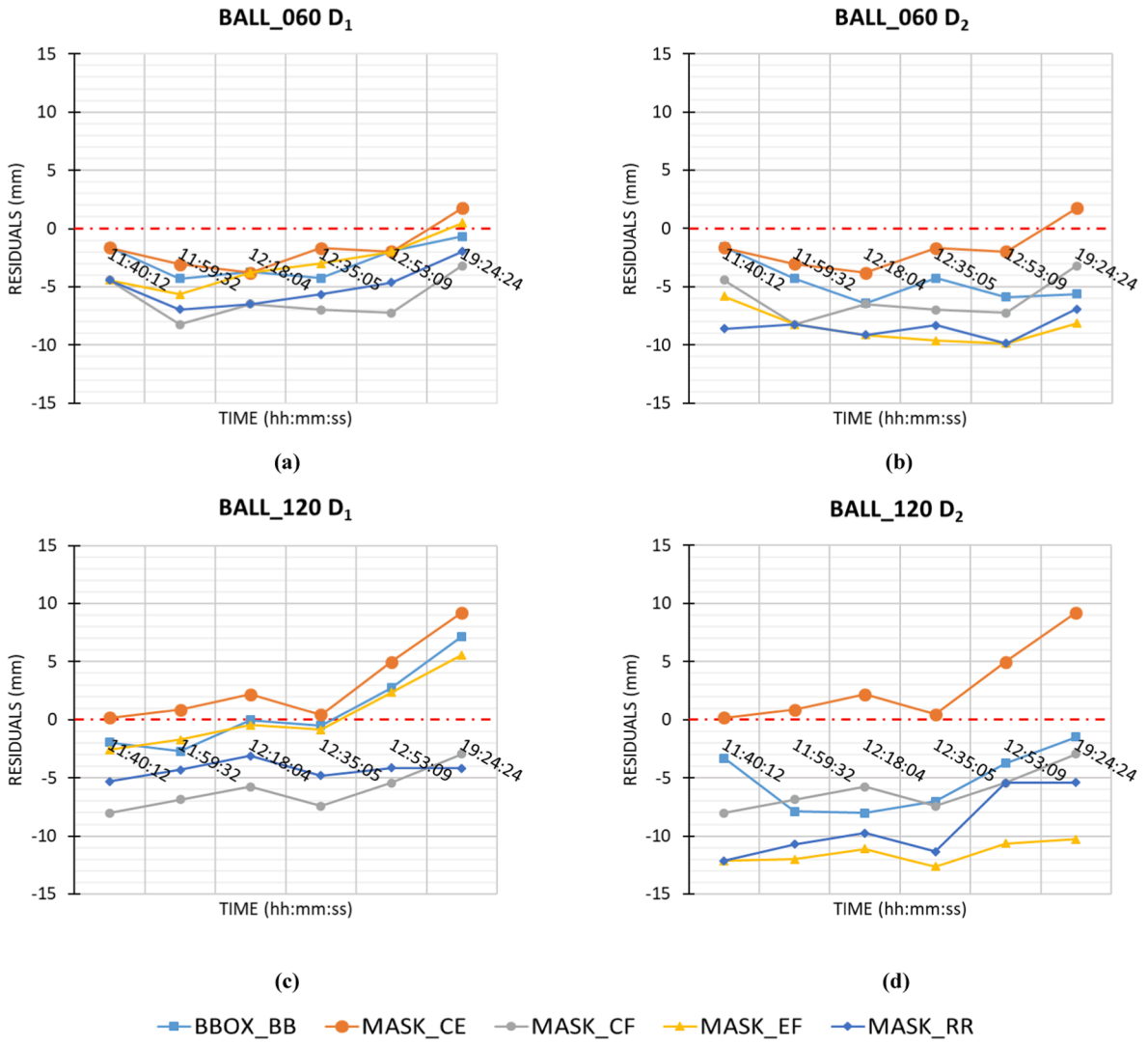


Fig. 10. Range of errors (estimated diameter – laboratory diameter) relating to the calibration spheres. a-b) BALL_060, laboratory diameter: $D_1 = 60.0$ mm, $D_2 = 60.0$ mm. c-d). BALL_120, laboratory diameter: $D_1 = 120.0$ mm, $D_2 = 120.0$ mm. The depth was estimated using the average (AVG) method.

3.3. Optimal algorithm for apple fruit sizing

Different methods for estimating the D_1 and D_2 axes were compared for the non-occluded apples in Table 4. The best estimates of the D_1 major axis were obtained when the ROI was identified using a binary mask (MASK), and then using rotate rectangles (RR). The algorithm was completed by converting the previous measurements to mm using the most repeated distance between the object and the camera (MOD depth). The mean distance (AVG) and the minimum distance (MIN) techniques are other options that could be considered. Errors of less than 5% (MAPE) were obtained when applying these sizing options, resulting in average deviations from real measurements of between 3 and 3.5 mm (MAE).

Different results were obtained when estimating the D_2 minor axis. The results obtained when using rotated rectangles were improved by fitting bounding boxes (BB, without masks) or even fitting ellipses (EF). However, in the latter case, the choice of the depth estimation method proved practically irrelevant. As before, errors (MAPE) were below 5%, giving mean deviations (MAE) of about 3 mm. The generalised use of the sizing algorithm could therefore be regarded as satisfactory, although attention should be paid to the use of different methods depending on whether the major or minor axis of the apples is estimated.

Another notable aspect was the poor performance when circles were used that enclosed the fruit region (CE method), with the apple sizing procedure producing the largest estimation errors for both the D_1 and D_2 axes (Table 4). In contrast to this trend, the rest of the methods seemed to show similar characteristics, at least when the maximum estimation error (MAPE) was set at 10%. This can be better appreciated through the visual interpretation of Fig. 12. In fact, adjusting the ROI using properly rotated rectangles (RR) resulted in good estimates of apple size using both axes, without the type of depth (mean, modal or minimum) seeming to have any significant influence. The use of bounding boxes (BB) was very close in performance (and even better for D_2). The results of using the circle fitting (CF) and the ellipses fitting (EF) methods were also noteworthy, producing somewhat larger errors, but without these exceeding 8% (Table 4 and Fig. 12).

In the case of the set of occluded apples (Table 5 and Fig. 13), the results varied considerably in terms of the recommended methods, in addition to producing worse estimates (MAPE always exceeded 5%). The use of bounding boxes (BB) was the most recommended option. Complemented with modal or mean depths, this method was at least able to keep the level of estimation errors (MAPE) below the 10% threshold for both D_1 and D_2. All the other methods failed in this regard, producing larger estimation errors (Fig. 13). In this set of methods, which were not as well-adapted to dealing with occluded apples, the circle fitting (CF) approach particularly stood out. This contrasted with

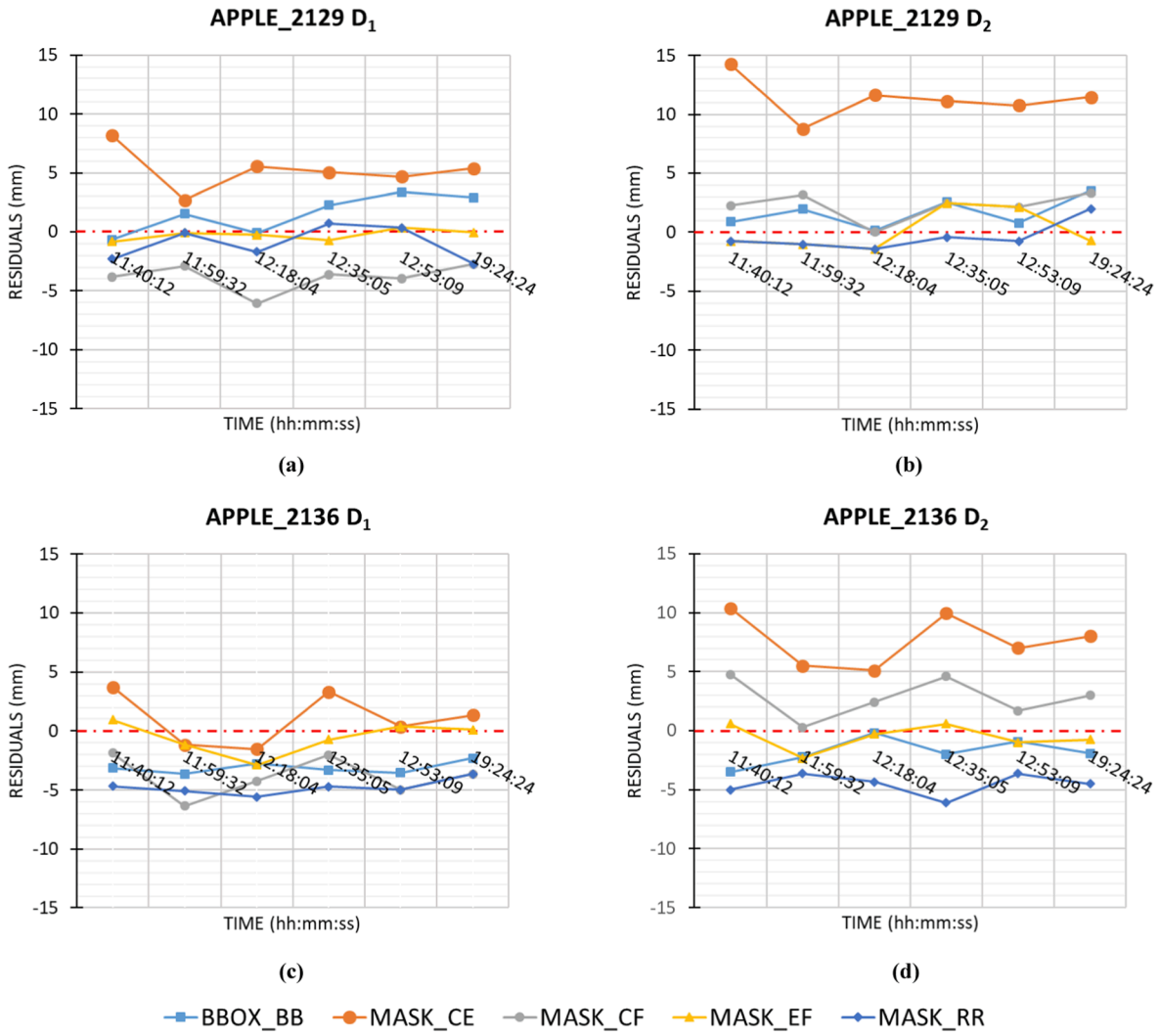


Fig. 11. Range of errors (estimated diameter – laboratory diameter) relating to the non-occluded apples. a-b) Apple #2129, laboratory diameter: $D_1 = 69.79$ mm, $D_2 = 75.87$ mm. c-d) Apple #2136, laboratory diameter: $D_1 = 66.27$ mm, $D_2 = 59.62$ mm. The depth was estimated using the average (AVG) method.

what happened with non-occluded apples. In the case of the occluded apples, circle enclosing (CE method) seemed to work better than methods that tried to fit rotated rectangles (RR) or ellipses (EF). As shown in Fig. 13, it was particularly difficult to estimate minor axis D_2 on occluded apples, while it was possible to use different methods interchangeably to estimate the major axis D_1. Whatever the case, depth approximation should be carried out using either the modal (MOD) or mean (AVG) method, and avoiding the calculation of the minimum depth in occluded apples.

3.4. Optimal combined sizing algorithm and allometric model for predicting apple weight

For non-occluded fruits (Table 6), sizing the apples using circle fitting (CF) and the subsequent application of the linear allometric model (3) (Table 3, Section 3.2) was found to be the algorithm option that provided the best weight predictions, with an error (MAPE) of only 5.1%. Very similar results, in terms of reliability, were obtained with options using ellipses (EF), or even bounding boxes (BB), as sizing methods before subsequently applying the same linear model (3); these approaches produced prediction errors of less than 6%. The results obtained with occluded apples were somewhat different (Table 6), with the best ranking corresponding to sizing with ellipses (EF) and making allometric weight predictions using model (1): a simple linear regression that only uses the measurement corresponding to the major axis D_1 of the apples as a predictor. As expected, the error (MAPE) subsequently increased to the very significant level of 18.3% (Table 6).

Somewhat surprisingly, the sizing methods that provided best results in terms of estimating apple size (Section 3.3), performed considerably less well when the allometric model was incorporated in order to predict the final weight. As allometric models are predictive, it is likely that some sort of compensation effect to ameliorate the prediction error would have occurred since estimated (rather than actual) measures of size were used as predictors in the models. A clear example of this can be seen in the case of non-occluded apples. While the use of rotated rectangles (RR) and bounding boxes (BB) seemed to be the best options for estimating D_1 and D_2 separately (Section 3.3), circle fitting (CF) proved the most recommendable sizing option as a first stage in the weight prediction algorithm. As shown in the previous section, other sizing approaches ranked better than the use of circle fitting (CF). Specifically, it was not among the best options for making estimates of D_2 (7% error).

To analyse the influence of allometric models on weight predictions, the best (first) weighting options for non-occluded and occluded apples were combined with all the different models listed in Table 3. The resulting prediction errors are shown in Table 7.

For weight predictions involving non-occluded apples, linear models were the best options, with prediction errors (MAPE) ranging from 5.1% (multiple linear model using D_1 and D_2 as predictors) to 8.3% (simple

Table 3
Allometric models used to predict apple fruit weight based on the major axis (D_1) and minor axis (D_2) geometric predictors of the fruit. The models were obtained from laboratory data.

Model identifier	*Linear models*	Goodness-of-fit (R^2)	Training dataset n = 568 (*RMSE*) [g]	Test dataset n = 489 (*RMSE*) [g]
(1)	$W = \beta_0 + \beta_1 D_1 + \varepsilon$ $\widehat{W} = -162.79 + 4.60 \times D_1$	0.942	14.98	15.93
(2)	$W = \beta_0 + \beta_1 D_1 + \beta_2 D_1^2 + \beta_3 D_1^3 + \beta_4 D_1^4 + \varepsilon$ $\widehat{W} = -298.4 + 25.47 \times D_1 - 0.78 \times D_1^2 + 0.01 \times D_1^3 - 0.000048 \times D_1^4$	0.979	8.97	9.29
(3)	$W = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \varepsilon$ $\widehat{W} = -161.64 + 2.48 \times D_1 + 2.22 \times D_2$	0.949	14.01	15.19
(4)	$W = \beta_0 + \beta_1 (D_1^2 D_2) + \varepsilon$ $\widehat{W} = 2.59 + 0.00046 \times D_1^2 D_2$	0.985	7.57	7.80
(5)	$W = \beta_0 + \beta_1 (D_1 D_2^2) + \varepsilon$ $\widehat{W} = 4.32 + 0.00048 \times D_1 D_2^2$	0.980	8.86	9.13
	Nonlinear models			
(6)	$W = \beta_0 \times D_1^{\beta_1} + \varepsilon$ $\widehat{W} = 0.00065 \times D_1^{2.91}$	0.989	9.36	9.32
(7)	$W = \beta_0 \times D_1^{\beta_1} \times D_2^{\beta_2} + \varepsilon$ $\widehat{W} = 0.00071 \times D_1^{1.80} \times D_2^{1.11}$	0.993	7.51	7.86

linear model using the combined predictor $D_1^2 D_2$). The use of nonlinear models increased the error (MAPE) to over 9%. The highly deviant polynomial model is an option that should be discarded when making this type of prediction. In fact, our research seemed to confirm that the use of polynomial models (such as Marini et al., 2019) is a good option when real fruit measurements are used as input variables. However, with uncertain values as input variable, a polynomial model may generate unacceptable errors.

4. Discussion

The main contribution of this work is the development of algorithms that can simultaneously predict apple fruit size and weight on the tree based on measurements taken using an RGB-D camera. However, it is known that RGB-D cameras do not tend to perform particularly well under conditions of direct sunlight. In this regard, Gené-Mola et al. (2020a) established 2000 lx as the illuminance threshold above which the performance of the Kinect v2 camera is adversely affected. In this work, the morning captures were registered with an illuminance of greater than 15,000 lx (Fig. 9), using an Azure Kinect camera. No significant differences were appreciated in size estimates compared to those registered in the late afternoon (500 lx) (Figs. 10 and 11). These results indicate that the Azure Kinect was not significantly influenced by

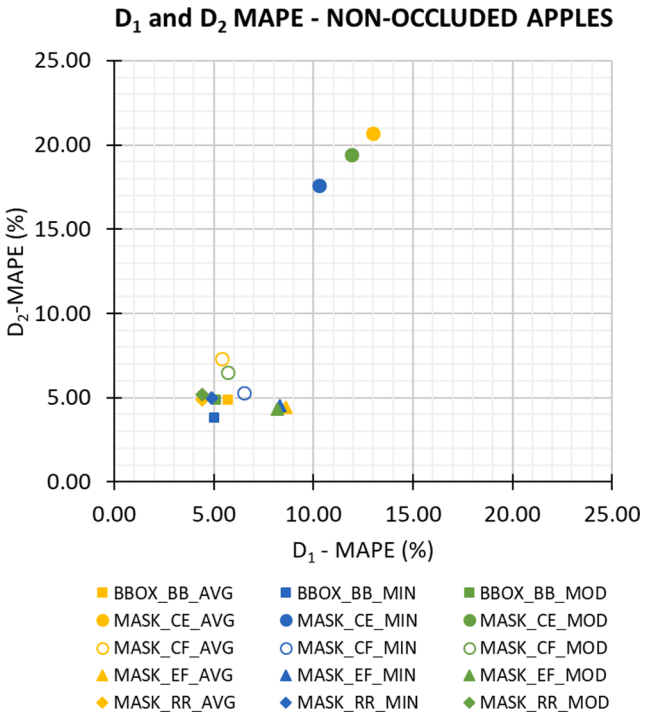


Fig. 12. Comparison of the Mean Absolute Percentage Error (MAPE) between D_1 and D_2 applied to the set of non-occluded apples (n = 9).

Table 4
Ranking of the methods applied to the non-occluded apple dataset (n = 9) organised according to major and minor diameter.

D_1						D_2					
Pixel sel.	**ROI**	**Depth**	**RMSE [mm]**	**MAE [mm]**	**MAPE [%]**	**Pixel sel.**	**ROI**	**Depth**	**RMSE [mm]**	**MAE [mm]**	**MAPE [%]**
MASK	RR	MOD	3.801	3.156	4.4	BBOX	BB	MIN	3.427	2.573	3.8
MASK	RR	AVG	3.859	3.201	4.4	MASK	EF	MOD	3.933	2.901	4.3
MASK	RR	MIN	4.277	3.527	4.9	MASK	EF	AVG	3.928	2.958	4.4
BBOX	BB	MIN	4.404	3.646	5.0	MASK	EF	MIN	4.048	3.047	4.5
BBOX	BB	MOD	4.606	3.729	5.1	BBOX	BB	AVG	4.104	3.302	4.9
MASK	CF	AVG	4.473	3.865	5.4	BBOX	BB	MOD	4.009	3.289	4.9
MASK	CF	MOD	4.706	4.122	5.7	MASK	RR	AVG	3.996	3.227	4.9
BBOX	BB	AVG	5.029	4.168	5.7	MASK	RR	MIN	4.203	3.291	5.0
MASK	CF	MIN	5.383	4.639	6.5	MASK	RR	MOD	4.177	3.453	5.2
MASK	EF	MOD	8.890	5.955	8.2	MASK	CF	MIN	4.427	3.638	5.3
MASK	EF	MIN	8.562	6.047	8.3	MASK	CF	MOD	5.106	4.451	6.5
MASK	EF	AVG	9.295	6.266	8.6	MASK	CF	AVG	5.678	4.975	7.3
MASK	CE	MIN	10.535	7.545	10.3	MASK	CE	MIN	13.660	12.072	17.6
MASK	CE	MOD	11.511	8.695	11.9	MASK	CE	MOD	14.966	13.286	19.4
MASK	CE	AVG	12.115	9.489	13.0	MASK	CE	AVG	15.739	14.131	20.7

D_1 = Major Diameter. D_2 = Minor Diameter. BBOX = Bounding Box. MASK = Mask. BB = Bounding Box. RR = Rotated Rectangle. EF = Ellipse Fitting. CE = Circle Enclosing. CF = Circle Fitting. AVG = Average depth. MOD = Modal depth. MIN = Minimum depth.

Table 5
Ranking of the methods applied to the occluded apple dataset (n = 17), organised by major and minor diameter.

D_1						D_2					
Pixel sel.	**ROI**	**Depth**	**RMSE [mm]**	**MAE [mm]**	**MAPE [%]**	**Pixel sel.**	**ROI**	**Depth**	**RMSE [mm]**	**MAE [mm]**	**MAPE [%]**
BBOX	BB	MOD	7.214	6.177	8.8	BBOX	BB	MOD	6.612	5.652	8.1
MASK	EF	MOD	8.033	6.558	9.2	BBOX	BB	AVG	7.248	6.170	8.8
MASK	EF	AVG	8.169	6.694	9.4	MASK	CE	MIN	9.664	7.714	11.4
BBOX	BB	AVG	7.693	6.707	9.5	MASK	CE	MOD	10.144	8.379	12.4
MASK	RR	MOD	8.417	7.015	10.0	MASK	CE	AVG	10.247	8.428	12.5
MASK	CE	MIN	9.328	7.220	10.0	BBOX	BB	MIN	11.343	9.173	12.9
MASK	CE	MOD	8.794	7.328	10.2	MASK	CF	MOD	13.271	10.925	15.6
MASK	RR	AVG	8.570	7.263	10.3	MASK	CF	AVG	13.580	11.162	15.9
MASK	CE	AVG	8.866	7.405	10.3	MASK	RR	MOD	12.723	11.075	16.3
MASK	EF	MIN	10.492	8.424	11.7	MASK	RR	AVG	12.959	11.260	16.6
BBOX	BB	MIN	10.843	8.801	12.4	MASK	EF	MOD	14.502	12.614	18.4
MASK	RR	MIN	11.841	9.916	13.8	MASK	EF	AVG	14.724	12.763	18.6
MASK	CF	MOD	16.149	14.215	19.7	MASK	CF	MIN	16.956	14.206	20.2
MASK	CF	AVG	16.381	14.397	19.9	MASK	RR	MIN	16.449	14.448	21.0
MASK	CF	MIN	19.665	17.567	24.3	MASK	EF	MIN	17.998	16.043	23.2

D_1 = Major Diameter. D_2 = Minor Diameter. BBOX = Bounding Box. MASK = Mask. BB = Bounding Box. RR = Rotated Rectangle. EF = Ellipse Fitting. CE = Circle Enclosing. CF = Circle Fitting. AVG = Average depth. MOD = Modal depth. MIN = Minimum depth.

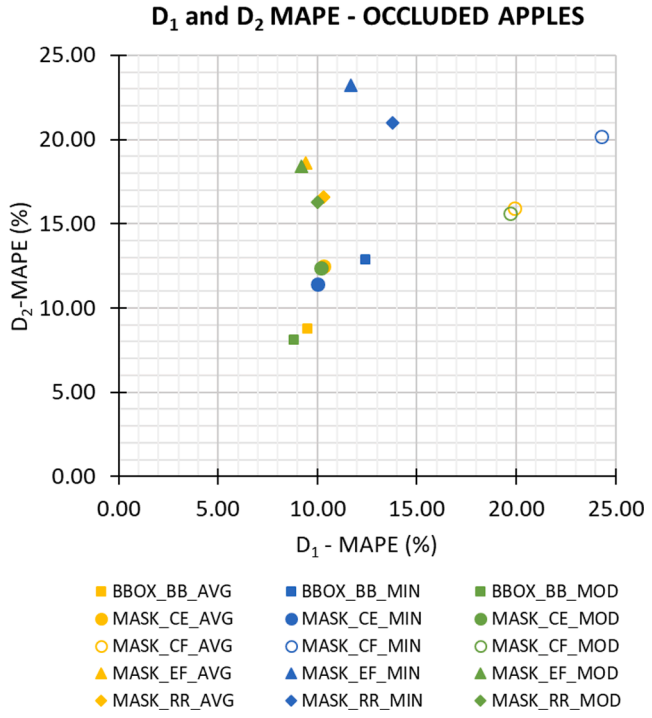


Fig. 13. Comparison of the Mean Absolute Percentage Error (MAPE) between D_1 and D_2 applied to the set of occluded apples (n = 17).

Table 6
Ranking of the methods applied to the non-occluded and occluded apple datasets for measurements of weight using average depth.

Weight predicted					
Pixel sel.	**ROI**	**Allometric weight prediction model**	**RMSE [g]**	**MAE [g]**	**MAPE [%]**
Non-occluded apple dataset (n = 9)					
MASK	CF	(3)	11.088	9.184	5.1
MASK	EF	(3)	12.244	10.100	5.6
BBOX	BB	(3)	13.019	10.121	5.7
MASK	CF	(1)	15.946	12.829	7.0
MASK	RR	(1)	17.970	14.714	8.1
MASK	RR	(3)	17.481	14.785	8.1
BBOX	BB	(1)	18.374	14.646	8.2
MASK	CF	(4)	17.901	14.237	8.3
MASK	EF	(5)	20.200	15.101	8.6
BBOX	BB	(5)	20.821	16.090	9.0
Occluded apple dataset (n = 17)					
MASK	EF	(1)	39.584	31.608	18.3
BBOX	BB	(3)	42.489	34.052	18.6
MASK	CE	(1)	36.419	29.878	18.8
BBOX	BB	(1)	40.913	33.311	18.9
MASK	CE	(3)	39.209	32.116	20.6
MASK	RR	(1)	47.047	38.802	21.9
BBOX	BB	(7)	46.288	39.197	22.9
BBOX	BB	(5)	49.471	40.881	23.2
MASK	EF	(6)	50.631	40.775	23.5
BBOX	BB	(4)	47.627	40.356	23.6

BBOX = Bounding Box. MASK = Mask. BB = Bounding Box. RR = Rotated Rectangle. EF = Ellipse Fitting. CE = Circle Enclosing. CF = Circle Fitting. Weight prediction model identifiers from Table 3.

sunlight, confirming findings reported by Neupane et al., (2021), who recommended the use of the Azure Kinect based on its robustness under direct sunlight and orchard conditions.

The size estimates for non-occluded and occluded apples are presented in Section 3.3. The estimation errors for non-occluded apples (Table 4: MAPE < 5 %; MAE = 3–3.5 mm; RMSE < 4 mm) were similar to those obtained in other studies using 3D sensing techniques such as LiDAR (MAE = 3.5–12.4 mm) (Tsoulias et al., 2020) or structure-from-motion photogrammetry (MAE = 3.7 mm) (Gené-Mola et al., 2021). These results were also comparable with those obtained using other RGB-D cameras on mango (RMSE = 4.3–4.9 mm) (Wang et al., 2017) and pomegranate (RMSE = 2.35 mm) (Yu et al., 2022) crops. As expected, the greatest errors were found when assessing occluded apples (Table 5: MAPE < 10 %; MAE = 6–8 mm; RMSE < 8 mm). The application of amodal instance segmentation to reconstruct the shape of occluded apples may, however, offer a way to improve these results (Gené-Mola et al., 2023).

Regarding fruit weight (Section 3.4), accurate estimates were obtained for non-occluded apples (Table 6, MAPE < 6 %) which were below the threshold of 10 % relative error usually accepted for harvest predictions (Uribeetxebarria et al., 2019). For occluded apples, the errors (MAPE) exceeded 18 % (Table 6) as the size estimates were less accurate. Given this result, one could consider the possibility of discarding readings for occluded apples, which tend to undermine yield predictions, as similar to what Neupane et al. (2022) pose in mango. When selecting the most appropriate methodology, a number of practical implementation issues need to be addressed, as well as the estimation errors. In this sense, the bounding box (BB) method has certain advantages over mask-based methods (CF, circle fitting; EF, ellipse fitting; CE, circle enclosing; RR, rotated rectangle), particularly in terms

Table 7
Ranking of allometric models (identifier in parentheses) once the best combined sizing-weighting option using average depth is selected for the non-occluded and occluded apple datasets.

Weight predicted					
Pixel sel.	**ROI**	**Allometric weight prediction model**	**RMSE [g]**	**MAE [g]**	**MAPE [%]**
Non-occluded apple dataset (n = 9)					
MASK	CF	(3)	11.088	9.184	5.1
		(1)	15.946	12.829	7.0
		(4)	17.901	14.237	8.3
		(7)	19.164	16.171	9.1
		(6)	18.938	15.143	9.3
		(5)	21.162	17.926	9.9
		(2)	257.367	251.633	138.4
Occluded apple dataset (n = 17)					
MASK	EF	(1)	39.584	31.608	18.3
		(6)	50.631	40.775	23.5
		(3)	53.724	45.352	25.6
		(7)	56.481	48.350	27.8
		(4)	56.776	48.523	27.9
		(5)	72.018	61.306	34.8
		(2)	271.061	255.175	140.9

MASK = Mask. EF = Ellipse Fitting. CF = Circle Fitting. Weight prediction model identifiers from Table 3.

of lower computational cost and more direct integration with current object detectors.

In the case of allometric models (Table 7), even with good results from the linear model (3), $W = \beta_0 + \beta_1 D_1 + \beta_2 D_2$, a degree of caution is required given that problems of multicollinearity may make it preferable to use other, more stable, single-predictor models. Models (1), $W = \beta_0 + \beta_1 D_1$, and (4), $W = \beta_0 + \beta_1 (D_1^2 D_2)$, are therefore strong candidates for use, rather than the aforementioned multiple model.

The compensatory effect that seemed to occur between the sizing algorithms and the allometric models should be viewed with caution. It may entail certain problems, but these are inherent to the sequential use of sizing algorithms obtained via machine vision and properly tested allometric models. In non-occluded apples, good RR (rotated rectangle)- and BB (bounding box)-based sizing algorithms continue to be valid options when their outputs are implemented in the appropriate allometric models (Table 6). Although the MAPE increased when delimiting non-occluded apples using the CF (circle fitting) algorithm, in no case was the threshold value of 10% exceeded. In general, sizing algorithms that achieve apple size estimation errors (MAPE) of below 10% (Table 4; Fig. 12) are also valid options and can complement the allometric model and predict yield in an acceptable way (Table 6). More specifically, with a MAPE of $< 8.1\%$, the results of our research would suggest that any of the algorithms (BB, bounding box; RR, rotated rectangle; EF, ellipse fitting; CF, circle fitting) could be applied when entering estimates of the major and minor axis of apples in a linear allometric model.

5. Conclusions

Time-of-flight RGB-D cameras offer a good option for sizing apples using computer vision algorithms for subsequent weight predictions made with appropriate allometric models. More specifically, the Azure Kinect camera is a relatively cheap device that performs well in agricultural environments under variable lighting conditions throughout the day.

The sizing methods that should be applied will differ depending on whether apples are non-occluded or occluded. The MAPE value was generally below 5 % for non-occluded apples (after adjusting their shape using rotated rectangles), while it increased to almost 10 % in occluded apples (adjusting the shape using bounding boxes). These sizing results were similar to those obtained with other techniques (e.g. LiDAR, structure-from-motion) but can be achieved using an affordable RGB-D camera with a low computational cost. In the case of depth measurements, for the final millimetric sizing of apples, average depths and modal values are equally recommendable options. When expanding the goal to weight prediction, in non-occluded apples, the rotated rectangles method should be replaced by fitting circles, ellipses or even bounding boxes, to then complement the sizing algorithm with a linear allometric model that uses both the major and minor axes as predictors. When fitting circles, the final MAPE (for weight prediction) was only 5.1 %. A non-additive error effect (or compensation) therefore occurs, despite the fact that sizing using circles (with a MAPE of 5.4 % on the major axis and of 7.3 % on the minor axis) and allometric modelling were implemented sequentially.

These promising sizing and weight prediction results open up the possibility of using RGB-D cameras for real-time fruit orchard characterization. Future work will include the implementation of an appropriate object detector to complete the acquisition-processing-yield prediction cycle.

CRediT authorship contribution statement

Juan C. Miranda: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Resources, Data curation, Writing – original draft, Writing – review & editing. **Jaume Arnó:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Jordi Gené-Mola:** Resources, Data curation, Writing – original draft, Writing – review & editing. **Jaume Lordan:** Resources, Data curation, Writing – original draft, Writing – review & editing. **Luis Asín:** Resources, Data curation, Writing – original draft, Writing – review & editing. **Eduard Gregorio:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was partly funded by the Department of Research and Universities of the Generalitat de Catalunya (grants 2017, SGR 646 and 2021 LLAV 00088), by the Spanish Ministry of Science and Innovation / AEI/10.13039/501100011033 / ERDF (grants RTI2018-094222-B-I00 [PAgFRUIT project], PID2021-126648OB-I00 [PAgPROTECT project]) and by the Spanish Ministry of Science and Innovation / AEI/10.13039/501100011033 / European Union NextGeneration / PRTR (grant-TED2021-131871B-I00 [DIGIFRUIT project]). We would also like to thank the Secretariat of Universities and Research of the Department of Business and Knowledge of the Generalitat de Catalunya and the European Social Fund (ESF) for financing Juan Carlos Miranda's pre-doctoral fellowship (2020 FI_B 00586). The work of Jordi Gené-Mola was supported by the Spanish Ministry of Universities through a Margarita Salas postdoctoral grant funded by the European Union - NextGenerationEU.

References

Alibabaei, K., Gaspar, P.D., Lima, T.M., Campos, R.M., Girão, I., Monteiro, J., Lopes, C. M., 2022. A review of the challenges of using deep learning algorithms to support

decision-making in agricultural activities. Remote Sens. 14, 638. https://doi.org/10.3390/rs14030638.
Anderson, N.T., Walsh, K.B., Wulfsohn, D., 2021. Technologies for forecasting tree fruit load and harvest timing - from ground, sky and time. Agronomy 11 (7), 1409. https://doi.org/10.3390/agronomy11071409.
Bargoti, S., 2016. Pychet Labeller - An object annotation toolbox [WWW Document]. URL https://github.com/acfr/pychetlabeller (accessed 7.25.23).
Dalmases, J., Pascual, M., Urbina, V., Blanco, R., 1998. Allometric relationships in peach fruit. Acta Hortic. 465, 415–424. https://doi.org/10.17660/ActaHortic.1998.465.52.
Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
Faraway, J.J., 2016. Linear Models with R. Chapman and Hall/CRC, New York, NY. https://doi.org/10.1201/b17144.
Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., Zhang, Q., 2020. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. Comput. Electron. Agric. 177, 105687 https://doi.org/10.1016/j.compag.2020.105687.
Gené-Mola, J., Llorens, J., Rosell-Polo, J.R., Gregorio, E., Arnó, J., Solanelles, F., Martínez-Casasnovas, J.A., Escolà, A., 2020a. Assessing the performance of rgb-d sensors for 3d fruit crop canopy characterization under different operating and lighting conditions. Sensors (Switzerland) 20 (24), 1–21. https://doi.org/10.3390/s20247072.
Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J.R., Morros, J.-R., Ruiz-Hidalgo, J., Vilaplana, V., Gregorio, E., 2020b. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. Comput. Electron. Agric. 169, 105165 https://doi.org/10.1016/j.compag.2019.105165.
Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J.R., Escolà, A., Gregorio, E., 2021. In-field apple size estimation using photogrammetry-derived 3D point clouds: Comparison of 4 different methods considering fruit occlusions. Comput. Electron. Agric. 188, 106343 https://doi.org/10.1016/j.compag.2021.106343.
Gené-Mola, J., Ferrer-Ferrer, M., Gregorio, E., Blok, P.M., Hemming, J., Morros, J.-R., Rosell-Polo, J.R., Vilaplana, V., Ruiz-Hidalgo, J., 2023. Looking behind occlusions: a study on amodal segmentation for robust on-tree apple fruit size estimation. Comput. Electron. Agric. 209, 107854 https://doi.org/10.1016/j.compag.2023.107854.
Gongal, A., Karkee, M., Amatya, S., 2018. Apple fruit size estimation using a 3D machine vision system. Inf. Process. Agric. 5 (4), 498–503. https://doi.org/10.1016/j.inpa.2018.06.002.
Gregorio, E., Llorens, J., 2021. Sensing crop geometry and structure. In: Kerry, R., Escolà, A. (Eds.), Sensing Approaches for Precision Agriculture. Springer International Publishing, Cham, pp. 59–92. https://doi.org/10.1007/978-3-030-78431-7_3.
Hacking, C., Poona, N., Manzan, N., Poblete-Echeverría, C., 2019. Investigating 2-D and 3-D proximal remote sensing techniques for vineyard yield estimation. Sensors 19 (17), 3652. https://doi.org/10.3390/s19173652.
He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., Dhupia, J., 2022. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. Comput. Electron. Agric. 195, 106812 https://doi.org/10.1016/j.compag.2022.106812.
Khoshnam, F., Tabatabaeefar, A., Varnamkhasti, M.G., Borghei, A., 2007. Mass modeling of pomegranate (Punica granatum L.) fruit with some physical characteristics. Sci. Hortic. 114 (1), 21–26. https://doi.org/10.1016/j.scienta.2007.05.008.
Kurtser, P., Ringdahl, O., Rotstein, N., Berenstein, R., Edan, Y., 2020. In-field grape cluster size assessment for vine yield estimation using a mobile robot and a consumer level RGB-D camera. IEEE Robot. Autom. Lett. 5 (2), 2031–2038. https://doi.org/10.1109/LRA.2020.2970654.
Lakso, A.N., Corelli Grappadelli, L., Barnard, J., Goffinet, M.C., 1995. An expolinear model of the growth pattern of the apple fruit. J. Hortic. Sci. 70 (4), 389–394. https://doi.org/10.1080/14620316.1995.11515308.
Marini, R.P., Schupp, J.R., Baugher, T.A., Crassweller, R., 2019. Relationships between fruit weight and diameter at 60 days after bloom and at harvest for three apple cultivars. HortScience 54 (1), 86–91. https://doi.org/10.21273/HORTSCI13591-18.
Meier, U., 2018. Growth stages of mono- and dicotyledonous plants: BBCH Monograph. https://doi.org/10.5073/20180906-074619.
Mengoli, D., Bortolotti, G., Piani, M., Manfrini, L., 2022. On-line real-time fruit size estimation using a depth-camera sensor, in: 2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor). IEEE, pp. 86–90. https://doi.org/10.1109/MetroAgriFor55389.2022.9964960.
Microsoft, 2022. Azure Kinect DK hardware specifications [WWW Document]. URL https://learn.microsoft.com/en-us/azure/kinect-dk/hardware-specification (accessed 9.19.23).
Miranda, J.C., Arnó, J., Gené-Mola, J., Fountas, S., Gregorio, E., 2023. AKFruitYield: AK_SW_BENCHMARKER - Azure Kinect Size Estimation & Weight Prediction Benchmarker [WWW Document]. URL https://pypi.org/project/ak-sw-benchmarker/ (accessed 9.19.23).
Miranda, J.C., Gené-Mola, J., Arnó, J., Gregorio, E., 2022. AKFruitData: A dual software application for Azure Kinect cameras to acquire and extract informative data in yield tests performed in fruit orchard environments. SoftwareX 20, 101231. https://doi.org/10.1016/j.softx.2022.101231.
Mitchell, P.D., 1986. Pear fruit growth and the use of diameter to estimate fruit volume and weight. HortScience 21 (4), 1003–1005. https://doi.org/10.21273/HORTSCI.21.4.1003.
Neupane, C., Pereira, M., Koirala, A., Walsh, K.B., 2023. Fruit sizing in orchard: a review from caliper to machine vision with deep learning. Sensors 23, 3868. https://doi.org/10.3390/s23083868.
Neupane, C., Koirala, A., Wang, Z., Walsh, K.B., 2021. Evaluation of depth cameras for use in fruit localization and sizing: finding a successor to Kinect v2. Agronomy 11 (9), 1780. https://doi.org/10.3390/agronomy11091780.
Neupane, C., Koirala, A., Walsh, K.B., 2022. In-Orchard sizing of mango fruit: 1. comparison of machine vision based methods for on-the-go estimation. Horticulturae 8, 1223. https://doi.org/10.3390/horticulturae8121223.
Rosell-Polo, J.R., Cheeinx, F.A., Gregorio, E., Andújar, D., Puigdomènech, L., Masip, J., Escolà, A., 2015. Advances in structured light sensors applications in precision agriculture and livestock farming. Adv. Agron. 133, 71–112. https://doi.org/10.1016/bs.agron.2015.05.002.
Spreer, W., Müller, J., 2011. Estimating the mass of mango fruit (Mangifera indica, cv. Chok Anan) from its geometric dimensions by optical measurement. Comput. Electron. Agric. 75 (1), 125–131. https://doi.org/10.1016/j.compag.2010.10.007.
Stajnko, D., Rozman, Č., Pavlovič, M., Beber, M., Zadravec, P., 2013. Modeling of "Gala" apple fruits diameter for improving the accuracy of early yield prediction. Sci. Hortic. (Amsterdam) 160, 306–312. https://doi.org/10.1016/j.scienta.2013.06.003.
Tabatabaeefar, A., Rajabipour, A., 2005. Modeling the mass of apples by geometrical attributes. Sci. Hortic. (Amsterdam) 105 (3), 373–382. https://doi.org/10.1016/j.scienta.2005.01.030.
Tsoulias, N., Paraforos, D.S., Xanthopoulos, G., Zude-Sasse, M., 2020. Apple shape detection based on geometric and radiometric features using a LiDAR laser scanner. Remote Sens. 12 (15), 2481. https://doi.org/10.3390/rs12152481.
Uribeetxebarria, A., Martínez-Casasnovas, J.A., Tisseyre, B., Guillaume, S., Escolà, A., Rosell-Polo, J.R., Arnó, J., 2019. Assessing ranked set sampling and ancillary data to improve fruit load estimates in peach orchards. Comput. Electron. Agric. 164, 104931 https://doi.org/10.1016/j.compag.2019.104931.
Wang, D., Li, C., Song, H., Xiong, H., Liu, C., He, D., 2020. Deep learning approach for apple edge detection to remotely monitor apple growth in orchards. IEEE Access 8, 26911–26925. https://doi.org/10.1109/ACCESS.2020.2971524.
Wang, Z., Walsh, K.B., Verma, B., 2017. On-tree mango fruit size estimation using RGB-D images. Sensors (Switzerland) 17 (12), 1–15. https://doi.org/10.3390/s17122738.
Welte, H.F., 1990. Forecasting harvest fruit size during he growing season. Acta Hortic. 276, 275–282. https://doi.org/10.17660/ActaHortic.1990.276.32.
Yu, T., Hu, C., Xie, Y., Liu, J., Li, P., 2022. Mature pomegranate fruit detection and location combining improved F-PointNet with 3D point cloud clustering in orchard. Comput. Electron. Agric. 200, 107233 https://doi.org/10.1016/j.compag.2022.107233.