Universitat de Lleida

Open source software and benchmarking of computer vision algorithms for apple fruit detection, fruit sizing and yield prediction using RGB-D cameras

Juan Carlos Miranda

http://hdl.handle.net/10803/690455

PhD Thesis

Open source software and benchmarking of computer vision algorithms for apple fruit detection, fruit sizing and yield prediction using RGB-D cameras

Juan Carlos Miranda

A Thesis presented for the degree of

Doctor by the University of Lleida

March 2024

Universitat de Lleida

TESI DOCTORAL

Open source software and benchmarking of computer vision algorithms for apple fruit detection, fruit sizing and yield prediction using RGB-D cameras

Juan Carlos Miranda

Memòria presentada per optar al grau de Doctor
per la Universitat de Lleida
Programa de Doctorat en Ciència i Tecnologia Agrària i Alimentària

Directores
Eduard Gregorio López
Jaume Arnó Satorra

2024

- ¿No tienes miedo?

- Siempre tengo miedo, pero si no lo enfrento nunca avanzaré. El mundo es de los que se animan y quieren más.

☺

Acknowledgements

First of all, I thank Sandra (my wife), her family and mine, for their unconditional support during this journey. They have had a lot of patience and an ocean full of difficulties in the middle, the goal has been met and this stage has come to an end.

I thank Eduard Gregorio and Jaume Arnó, who have invested a lot of time in the development and correction of this thesis, and have provided great support throughout these years. Jaume's practicality and capacity for synthesis, together with Eduard's attention to detail, have allowed this research to be possible. Also, I would like to extend my thanks to the Research Group on AgroTIC and Precision Agriculture (GRAP) and all its members for hosting me during the time I have lived in Lleida. Thank you Alex Escolá for integrating me into the teaching staff.

My thanks to Professor Spyros Fountas, for having accepted me as a visitor within his research group. My stay in Athens has been fantastic thanks to the support of the members of the Smart Farming Technology Group (SFTG) of the Agricultural University of Athens (AUA), who integrated me from the beginning.

Agradecimientos

En primer lugar, agradezco a Sandra (mi esposa), a su familia y a la mía, por el apoyo incondicional durante este viaje. Han tenido mucha paciencia y un océano cargado de dificultades en medio, se ha cumplido la meta y esta etapa ha llegado a su fin.

Agradezco a Eduard Gregorio y Jaume Arnó, quienes han invertido mucho tiempo en el desarrollo y corrección de esta tesis, han brindado un gran acompañamiento a lo largo de estos años. La practicidad y capacidad de síntesis de Jaume junto al detallismo de Eduard, han permitido que esta investigación sea posible. También, quiero extiendo mis agradecimientos al Grupo de Investigación en AgróTICa y Agricultura de Precisión (GRAP) y a todos sus miembros por acogerme durante el tiempo que he vivido en Lleida. Gracias Alex Escolá por integrarme en el plantel de profesores.

Mis agradecimientos al profesor Spyros Fountas, por haberme aceptado como visitante dentro de su grupo de investigación. Mi estadía en Atenas ha sido fantástica gracias al acompañamiento de los miembros del Smart Farming Technology Group (SFTG) de la Agricultural University of Athens (AUA), quienes me integraron desde un primer momento.

Agraïments

En primer lloc, agraeixo a Sandra (la meva dona), a la seva família i a la meva, pel suport incondicional durant aquest viatge. Han tingut molta paciència i un oceà carregat de dificultats al mig, s'ha complert la meta i aquesta etapa ha arribat al final.

Agraeixo a Eduard Gregorio i Jaume Arnó, que han invertit molt de temps en el desenvolupament i la correcció d'aquesta tesi, han brindat un gran acompanyament al llarg d'aquests anys. La practicitat i la capacitat de síntesi de Jaume juntament amb el detallisme d'Eduard, han permès que aquesta investigació sigui possible. També vull estendre els meus agraïments al Grup de Recerca en AgròTICa i Agricultura de Precisió (GRAP) i a tots els seus membres per acollir-me durant el temps que he viscut a Lleida. Gràcies Alex Escolá per integrar-me al planter de professors.

Els meus agraïments al professor Spyros Fountas, per haver-me acceptat com a visitant dins del seu grup de recerca. La meva estada a Atenes ha estat fantàstica gràcies a l'acompanyament dels membres de l'Smart Farming Technology Group (SFTG) de l'Agricultural University of Athens (AUA), que em van integrar des d'un primer moment.

Institutional Acknowledgements

I would like to thank the University of Lleida for allowing me to be part of the staff as research teaching staff (PDI). In addition, this thesis would not have been possible without the financial support of the Generalitat de Catalunya and the European Social Fund (ESF), which have funded the pre-doctoral grant 2020 FI_B 00586. Also, this work has been partly funded by the Department of Research and Universities of the Generalitat de Catalunya (grants 2017 SGR 646 and 2021 LLAV 00088) and by the Spanish Ministry of Science and Innovation / AEI/10.13039/501100011033 / FEDER (grants RTI2018-094222 -B-I00 [PAgFRUIT project], PID2021-126648OB-I00 [PAgPROTECT project] and TED2021-131871B-I00 [DIGIFRUIT project]).

I would like to mention the external collaborators, who also gave their time and ideas. Thanks are due to the intervening members of the Institute for Agri-Food Research and Technology (IRTA) and the Department of Horticultural Engineering of the Leibniz Institute for Agricultural Engineering and Bioeconomics (ATB) for their support.

Abstract

This thesis focuses on the detection (counting) of fruits and estimation of their size and weight in apple orchards through the application of computer vision techniques. This work seeks to provide fruit growers with advanced tools and methodologies to help them make accurate harvest yield predictions. Counting (quantifying) and locating fruits represent previous steps to achieve these predictions. By knowing this information, fruit growers can schedule in advance the required resources for harvest and post-harvest (labor, transportation, storage), design sales strategies and, ultimately, optimize the profitability of their farms. In addition, it is also essential to control fruit quality parameters such as size and weight, which have a great influence on the market price and decision making for canopy management. On the other hand, today, counting, locating and measuring the size of fruits are repetitive tasks that require trained labor and that can be affected by fatigue and the subjective criterion of workers. Therefore, the manual execution of these tasks in fruit orchards of several hectares is not feasible. These reasons largely explain the current need to develop automatic tools that allow accurate in-field fruit detection and sizing.

The main objective of this thesis is to explore the capacity of RGB-D sensors to estimate the size and weight of fruits in apple trees. The main body of this work is made up of four articles that deal in detail with various stages throughout the process: i) a review of the state of the art on fruit sizing using artificial intelligence techniques and its challenges in field conditions; ii) the development of software tools for data acquisition in fruit orchards; iii) the development of algorithms for estimating fruit size and weight using allometric models; and iv) an experimental field evaluation of the implemented algorithms, applying combinations of fruit sizing methods and allometric models for weight prediction. The results obtained presented errors (MAPE) of less than 5 % in the estimation of the size of non-occluded apples and less than 5.1% for the prediction of their weight. These results open the possibility of using affordable RGB-D cameras in the short term for real-time characterization of fruit plantations. Furthermore, as a result of this thesis, a set of open access software tools has been made available to the public for in-field fruit detection and estimate their size and weight. In conclusion, the thesis represents a contribution towards the development of affordable tools to facilitate decision-making and help optimize fruit orchard management.

Resumen

Esta tesis se enfoca en la detección (conteo) de frutos y estimación de su tamaño y peso en plantaciones de manzanos mediante la aplicación de técnicas de visión por ordenador. Este trabajo busca proporcionar a los fruticultores herramientas y metodologías avanzadas para ayudarles a realizar predicciones de cosecha precisas. Contar (cuantificar) y localizar frutos representan pasos previos para lograr dichas predicciones. Al conocer esta información, los fruticultores pueden programar con antelación los recursos necesarios para la cosecha y poscosecha (mano de obra, transporte, almacenamiento), diseñar estrategias de ventas y, en definitiva, optimizar la rentabilidad de sus explotaciones. Además, también es fundamental controlar parámetros de calidad de la fruta como su tamaño y peso, los cuales tienen una gran influencia en el precio de mercado y en la toma de decisiones para el manejo del dosel. Por otro lado, a día de hoy, contar, localizar y medir el tamaño de los frutos son tareas repetitivas que requieren de mano de obra capacitada y que pueden verse afectadas por el cansancio y el criterio subjetivo de los trabajadores. Por lo que, la ejecución manual de estas tareas en plantaciones frutales de varias hectáreas no es viable. Estas razones explican en gran medida la necesidad actual de desarrollar herramientas automáticas que permitan detectar y medir los frutos en campo con precisión.

El principal objetivo de esta tesis es explorar la capacidad de los sensores RGB-D para estimar el tamaño y el peso de los frutos en manzanos. El cuerpo principal de este trabajo lo constituyen cuatro artículos que tratan en detalle diversas etapas a lo largo del proceso: i) una revisión del estado del arte sobre dimensionamiento de frutos utilizando técnicas de inteligencia artificial y sus desafíos en condiciones de campo; ii) el desarrollo de herramientas software para adquisición de datos en plantaciones frutícolas; iii) el desarrollo de algoritmos para estimación de tamaño y peso de frutos mediante modelos alométricos; y iv) una evaluación experimental en campo de los algoritmos implementados, aplicando combinaciones de métodos de dimensionamiento y modelos alométricos para la predicción de peso. Los resultados obtenidos presentaron errores (MAPE) inferiores al 5 % en la estimación del tamaño de manzanas no ocluidas y menores al 5.1 % para la predicción de su peso. Estos resultados abren la posibilidad de utilizar a corto plazo cámaras RGB-D asequibles para la caracterización en tiempo real de plantaciones frutales. Además, como resultado de esta tesis se ha puesto a disposición pública un conjunto de herramientas software de libre acceso para detectar los frutos en campo y estimar su tamaño y peso. En conclusión, la tesis supone una contribución

hacia el desarrollo de herramientas asequibles que faciliten la toma de decisiones y ayuden a optimizar la gestión de las explotaciones frutícolas.

Resum

Aquesta tesi s'enfoca a la detecció (compteig) de fruits i estimació de la seva grandària i pes en plantacions de pomeres mitjançant l'aplicació de tècniques de visió per ordinador. Aquest treball busca proporcionar als fructicultors eines i metodologies avançades per ajudar-los a fer prediccions de collita precises. Comptar (quantificar) i localitzar fruits representen passos previs per assolir aquestes prediccions. En conèixer aquesta informació, els fructicultors poden programar amb antelació els recursos necessaris per a la collita i la postcollita (mà d'obra, transport, emmagatzematge), dissenyar estratègies de vendes i, en definitiva, optimitzar la rendibilitat de les seves explotacions. A més, també és fonamental controlar paràmetres de qualitat de la fruita com la mida i el pes, els quals tenen una gran influència en el preu de mercat i en la presa de decisions per al maneig del dosser. D'altra banda, avui en dia, comptar, localitzar i mesurar els fruits són tasques repetitives que requereixen mà d'obra capacitada i que es poden veure afectades pel cansament i el criteri subjectiu dels treballadors. Per tant, l'execució manual d'aquestes tasques en plantacions de fruiters de diverses hectàrees no és viable. Aquestes raons expliquen en gran mesura la necessitat actual de desenvolupar eines automàtiques que permetin detectar i mesurar els fruits al camp amb precisió.

L'objectiu principal d'aquesta tesi és explorar la capacitat dels sensors RGB-D per estimar la grandària i el pes dels fruits en pomeres. El cos principal d'aquest treball el constitueixen quatre articles que tracten detalladament diverses etapes al llarg del procés: i) una revisió de l'estat de l'art sobre dimensionament de fruits utilitzant tècniques d'intel·ligència artificial i els seus desafiaments en condicions de camp; ii) el desenvolupament d'eines programari per a l'adquisició de dades en plantacions fructícoles; iii) el desenvolupament d'algoritmes per a l'estimació de la grandària i pes dels fruits mitjançant models al·lomètrics; i iv) una avaluació experimental en camp dels algoritmes implementats, aplicant combinacions de mètodes de dimensionament i models al·lomètrics per a la predicció de pes. Els resultats obtinguts van presentar errors (MAPE) inferiors al 5 % en l'estimació de la grandària de pomes no closes i menors al 5,1 % per a la predicció del pes. Aquests resultats obren la possibilitat d'utilitzar a curt termini càmeres RGB-D assequibles per a la caracterització en temps real de plantacions fruiteres. A més, com a resultat d'aquesta tesi s'ha posat a disposició pública un conjunt de programari de lliure accés per detectar els fruits en camp i estimar-ne la mida i el pes. En conclusió, la tesi suposa una contribució cap al desenvolupament d'eines assequibles que facilitin la presa de decisions i ajudin a optimitzar la gestió de les explotacions fructícoles.

Contents

Acknowledgements .... *i*

Agradecimientos .... *iii*

Agraïments .... *v*

Institutional Acknowledgements .... *vii*

Abstract .... *xiii*

Resumen .... *xv*

Resum .... *xvii*

1 Introduction *1*

1.1 The challenge of fruit detection and yield prediction using computer vision **2**

1.2 Affordable optical sensors for fruit detection **3**

1.3 The need to develop reliable software to use RGB-D cameras in orchard environments **7**

1.4 Objectives **7**

1.5 Thesis structure **8**

1.6 References **9**

Chapter 2: Methodology *15*

2.1 Azure Kinect DK RGB-D Camera Basics **16**

2.2 Software design bases for the Azure Kinect camera for use in fruit detection, sizing and yield prediction **17**

2.3 Programming environment **18**

2.4 References **19**

Chapter 3: Fruit sizing using AI: a review of methods and challenges *21*

3.1 Introduction **23**

3.2 Fruit detection based on handcrafted features **26**

3.3 Fruit detection based on deep learning **30**

3.4 Fruit size and maturity estimation **41**

3.5 Discussion and future trends **46**

3.6 Conclusions **52**

3.7 Acknowledgements **53**

3.8 References **53**

Chapter 4: AKFruitData: a dual software application for Azure Kinect cameras to acquire and extract informative data in yield tests performed in fruit orchard environments *77*

4.1 Motivation and significance **80**

4.2 Software description **82**

4.3 Illustrative examples **87**

4.4 Impact **90**

4.5 Conclusions 90
4.6 Declaration of competing interest 91
4.7 Acknowledgements 91
4.8 References 91
Chapter 5: AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards 95
5.1 Motivation and significance 98
5.2 Software description 99
5.3 Illustrative examples 105
5.4 Impact 107
5.5 Conclusions 108
5.6 Declaration of competing interest 109
5.7 Acknowledgements 109
5.8 References 109
Chapter 6: Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples 113
6.1 Introduction 115
6.2 Materials and methods 117
6.3 Results 128
6.4 Discussion 139
6.5 Conclusions 141
6.6 Acknowledgements 142
6.7 References 142
Chapter 7: Discussion 147
7.1 Progress in fruit detection and sizing using AI 148
7.2 Software development for RGB-D cameras in fruit orchards 149
7.3 Fruit sizing estimation and weight prediction 150
7.4 Future works 152
7.5 References 154
Chapter 8: Conclusions 157
Chapter 9: List of contributions 161
9.1 Journal papers included in the thesis 162
9.2 Software in open-access repositories 162
9.3 Scientific foreign-exchange 162
9.4 Research project participations 163

1 Introduction

This work has been carried out within the framework of the **PAgFRUIT research project (RTI2018-094222-B-I00)** whose objective is the development and application of precision agriculture technologies to optimize canopy management and sustainable crop protection in fruit orchards. This thesis focuses on fruit detection (counting) and sizing in apple orchards by applying computer vision techniques. This research's ultimate goal is to provide farmers with advanced tools and methodologies to help them in performing accurate fruit yield predictions.

Counting (quantifying) and locating fruits represent previous steps to achieve accurate yield predictions. These predictions allow fruit growers to schedule in advance the necessary harvest and postharvest resources (labor, transportation, storage), design sales strategies and optimize orchard profitability. In addition to their number, it is also essential to monitor fruit quality parameters such as their size and weight, which have a great influence on the market price. The knowledge of the aforementioned parameters is also important for decision-making in crop load management strategies.

Nowadays, counting, locating and sizing fruits are repetitive tasks that require trained labor that can be affected by fatigue and by the subjective criteria of the workers. Furthermore, the manual execution of these tasks in fruit orchards of several hectares is not feasible. These reasons greatly explain the current need for developing automatic tools that allow the fruit to be accurately detected and measured.

1.1 The challenge of fruit detection and yield prediction using computer vision

Fruit detection deals on finding a region of interest (ROI) in a given image, point cloud, or other type of data, and classifying it as fruit or background based on a confidence metric. Fruit location consists of obtaining the position of a detected fruit in a local or global coordinate system. Implementation of automatic solutions for in-field fruit detection and location is affected by several challenging factors. These factors include changes of the fruit shape and colour over the ripening process, the appearance of fruit occlusions, variable lighting conditions, among others. Furthermore, depending on the type of sensor and platform (manual, fixed, terrestrial) the optimal measurement conditions could be different. It should also be noted that these automatic solutions must deal with outdoor and unstructured environments, which adds further difficulty. It is therefore evident that the detection, location and sizing of fruits is a complex task that requires detailed study and it is currently a topic of great interest for the scientific community.

Counting and locating fruits with low or acceptable errors is a necessary step for achieving reliable yield predictions. However, this is not an easy task due to the time and resources usually required. Furthermore, the question arises as to whether perform the yield prediction through a complete fruit counting through the plot or by applying sampling techniques to collect data from only specific points.

A complete count (sweep) is difficult to be manually performed, even more so given the trend towards ever-large plots. The automatic systems for fruit orchard monitoring must be robust, deal with occlusions and fruit clusters, doing so quickly and, at the same time, lowering the error rate in detection and estimation.

In practice, farmers and insurance companies have chosen to apply sampling techniques to obtain predictions (BOE, 2005). In this case, the challenge is to design an efficient sampling method to reliably estimate the fruit load and/or yield (a few weeks before harvest) using reduced sample sizes. In addition to being efficient, it should be able to guide the fruit grower in the optimal location of the sampling points (smart points or specific areas to be sampled). There is also the added difficulty of developing a sampling method within the tree compatible with the computer vision technology developed for fruit detection.

1.2 Affordable optical sensors for fruit detection

The sensors used in fruit detection can be classified according to the emission and reception of light, by their ability to measure depth values, and by the number of receivers they use. It is possible to classify them according to the reception and emission of light (passive, active) and group them by their ability to measure depth. Passive sensors are those that make use of the reflection of light from the environment, while active sensors are those that use their own light source to obtain information about the distance of objects.

1.2.1 Passive sensors

Monocular cameras based on red-green-blue (RGB) sensors have been used in a large number of works due to their ease of use, low cost and the variety of devices available on the market. Currently, two technologies are present on the market: the charge-coupled device (CCD) and the complementary metal-oxide-semiconductor (CMOS) (Sarkar and Theuwissen, 2013). Both make use of a two-dimensional array of photosensitive units (pixels), which collect light and transform it into electrical signals. The number of pixels present in a device is known as resolution, a parameter associated with its quality (Solomon and Breckon, 2011). In CCDs, the signals from the pixels produce a uniform output and are processed in a conversion node to electrical pulses. In CMOSs, the conversion of light to electrical signals is carried out internally in each pixel. The CCD has shown its worth in environments that require high quality images, but at the cost of a longer reading time. On the other hand, the CMOS provides a higher reading speed, a lower power consumption and a lower manufacturing cost (Pajares et al., 2016).

Several limitations affect these devices, including changes in lighting, reflections and shadows. In addition, these sensors alone do not return information on depth in the scene, an important feature in fruit detection and location or in fruit size estimation, among other applications (Zhao et al., 2016).

Over the years, due to the need for depth information, techniques based on the principle of triangulation have been introduced. This involves using trigonometric relationships between the internal parameters of the camera (focal length) and a point of known distance to determine the distance of another point in a scene (Szeliski, 2011). Techniques based on this principle include stereo vision, structure-from-motion (SfM) and multi-view stereo (MVS) (Furukawa and Hernández, 2015; Klette, 2014; Özyeşil et al., 2017).

Thermal cameras, which acquire images of the infrared radiation from objects, are another type of passive sensor used in fruit detection. They have been used in cases where it is possible to identify objects by their thermal inertia in relation to the environment that surrounds them (Narvaez et al., 2017). Atmospheric changes strongly influence the data obtained, meaning that constant calibration and correction is required (Li et al., 2014).

Spectral sensors have also been used in fruit detection. Their operation is based on measuring the reflectance of objects at different wavelengths. Each material reflects wavelengths with characteristic values that allow it to be differentiated from other materials (Chaudhuri and Kotwal, 2014). Two sensor types can be distinguished: multispectral (MS) and hyperspectral (HS). MS sensors capture scene information in a small number of, not necessarily contiguous, bands of the electromagnetic spectrum, while HS sensors record reflected radiation in multiple contiguous bands, thus providing the spectral signature of the studied object. The use of this type of sensor has been limited, largely due to their high cost, ignorance of the technology and reduced availability outside of scientific settings (Lu et al., 2020). In addition to the above, the acquisition of spectral images, their processing and analysis represent a challenge due to the large volume and high dimensionality of the data (Khanal et al., 2020).

1.2.2 Active sensors

Advances in technologies and lower costs have allowed devices to appear on the market that obtain scene depth values based on the active emission of light. Three types can be distinguished according to their operating principle: time-of-flight (ToF), active stereo vision (ASV) and structured light (SL). A ToF sensor is based on the emission of light (pulsed or continuous) and the determination of its round-trip distance. In the case of light pulses, this distance is obtained by measuring their time of flight. In the case of amplitude modulated continuous emission, the distance is determined from the phase shift between the emitted signal and the received signal (Corti et al., 2016). ASV sensors calculate depth using epipolar geometry, observing an artificially projected light pattern from multiple cameras (Vit and Shani, 2018). Finally, SL sensors combine the projection of known light patterns with the data obtained by a camera and measure depth by triangulation (Sarbolandi et al., 2015).

For the 3D mapping of spaces, various authors refer to the use of light detection and ranging (LiDAR) devices. These sensors work under the ToF principle, returning point clouds from the surrounding environment as information. LiDAR sensors can detect small objects in noisy environments and enjoy good sensitivity for fruit detection tasks, but they cost more than other sensors and are affected by dusty environments, fog and humidity (X. Liu et al., 2019; Rosell and Sanz, 2012; Zhang et al., 2020).

Red-green-blue-depth (RGB-D) sensors constitute an extension to RGB cameras and have proven useful in agricultural environments. Their operation, depending on the model, can be included under the ToF, ASV or SL principles. Their low cost compared to other technologies and the ability to return color images together with depth and infrared information at high acquisition rates (Fu et al., 2020; Gregorio and Llorens, 2021), makes them good candidates for adoption in the agricultural sector. As a disadvantage, RGB-D sensors are sensitive to lighting in field conditions (Gené-Mola et al., 2020). Table 1 presents a summary of fruit detection works with details related to the sensors and platforms used.

Table 1.1 - Most used sensors and technologies in fruit detection

Sensors/ Techniques	Output data	Lighting conditions	Platforms	Crops	References
RGB camera	Image	Daylight	Terrestrial platform	Mangoes	(Qureshi et al., 2017)
		Daylight	Manually	Tomatoes	(G. Liu et al., 2019)
		Daylight	Tripod	Citrus, tomatoes, pumpkin, bitter gourds, towel gourds, mangoes.	(Lin et al., 2020)
		Daylight	Manual	Citrus	(Dorj et al., 2017)
RGB camera/stereo vision	Image/depth	Daylight	Terrestrial platform	Apples	(Onishi et al., 2019; Si et al., 2015)
		Daylight	Terrestrial platform	Kiwis	(Williams et al., 2019)
RGB camera/SfM – MVS	Image/depth	Controlled	Manual	Maize, sugar beets, sunflowers	(Martinez-Guanter et al., 2019)
		Controlled	Manual	Tomatoes	(Rose et al., 2015)
		Daylight	Manual	Apples	(Dong et al., 2018; Häni et al., 2020)
		Daylight	Manual	Coffee	(Avendano et al., 2017)
		Daylight	Terrestrial platform	Mangoes	(Stein et al., 2016)
		Daylight	Manual	Oranges	(Liu et al., 2018)
		Night-time	Manual	Apples	(Liu et al., 2018)
Thermal camera	Image/thermal data	Daylight	Tripod	Apples	(J. Feng et al., 2019)
		Daylight	Tripod	Citrus	(Gan et al., 2020, 2018)
Multi-/hyperspectral camera	Image/spectral data	Controlled	Fixed	Apples	(J. Feng et al., 2019; Zhang et al., 2015)
		N/M	Manual	Tomatoes	(Q. Feng et al., 2019)
		Daylight	Manual	Passion fruits	(Tu et al., 2020)
		Daylight	Manual	Pomegranates	(Zhang et al., 2021)
		Daylight	Terrestrial platform	Apples	(Gené-Mola et al., 2019b)
RGB-D camera		Daylight	Terrestrial platform	Grapes	(Kurtser et al., 2020)
		Daylight	Terrestrial platform	Grapes	(Milella et al., 2019)
		Daylight	Tripod	Peppers	(Vitzrabin and Edan, 2016)
LiDAR sensor	3D point cloud	Daylight	Terrestrial platform	Apples	(Gené-Mola et al., 2019a; Tsoulias et al., 2020)
		Daylight	Terrestrial platform	Almonds	(Underwood et al., 2016)
		Daylight	Manual	Oranges	(Méndez et al., 2019)
		Night-time	Terrestrial platform	Peppers	(Eizentals and Oka, 2016)

* References. Manual = includes sticks, phones and cameras operated manually. Fixed = fixed system, Tripod = images taken with tripods. Terrestrial platform = robots and vehicles. Daylight = images in daytime conditions. Night-time = images in night conditions. Controlled = images in controlled environments. N/M = not mentioned

1.3 The need to develop reliable software to use RGB-D cameras in orchard environments

In sections 1.1 and 1.2, main challenges in fruit detection and yield prediction have been mentioned and promising optical sensors technologies to be applied in these taks have also been reviewed. However, the adoption of these new technologies by farmers is currently limited by their cost, the difficulty of use, and the lack of data management skills of most potential users. Therefore, there is a need to develop robust, easy-to-use hardware and software solutions that allow the processing of data collected in the orchard and convert it into value-added information to farmers.

RGB-D sensors are characterized by their low cost and by the multimodal data (RGB image, depth, intensity) they provide, which makes them good candidates for fruit detection and sizing applications. Within this family of sensors, the Azure Kinect camera (Microsoft, Redmond, WA, USA) is a time-of-flight device that has replaced the popular Kinect v2 camera also manufactured by Microsoft. As detailed in Chapter 2, the Azure Kinect has proven to offer good performance in outdoor scenarios and it is the RGB-D sensor mostly used in this research. However, up to now there was no a specific software for this sensor to facilitate data acquisition and processing focused on automatic yield prediction. Thus, it was required the use of scripts or different tools whose execution involved excessive steps for the end user. One of the objectives of this work is to partially cover this shortage, proposing open source software tools, which can be adapted even for RGB-D sensors from other manufacturers.

1.4 Objectives

The main objective of this thesis is to explore the ability of RGB-D sensors to estimate fruit size and weight on the tree. The research focuses on apple orchards, a crop on which the Research Group in AgroICT & Precision Agriculture (GRAP) has carried out preliminary work using other types of sensors.

Specific objectives are listed below:

i) Carry out an updated bibliographic review on automatic fruit dection and sizing techniques in the field using computer vision.
ii) Develop a software tool for the acquisition and extraction of data recorded with RGB-D sensors in field conditions.
iii) Develop a software tool for benchmarking fruit size estimation and weight prediction algorithms.
iv) Develop a computer vision-based software tool for automatic fruit detection and yield estimation in the field.

v) Experimentally asses in a fruit orchard the performance of the algorithms proposed and implemented in the previous specific objectives.

1.5 Thesis structure

his doctoral thesis is presented as a compendium of articles following the Academic Regulations for Doctoral Courses at the University of Lleida, where each chapter can be considered a self-contained unit. Fig. 1.1 shows the structure of the thesis, indicating the published articles and the objectives with which they are aligned. Chapter 1 presents introductory aspects, challenges and related technologies. In Chapter 2, the applied methodology is briefly described. Chapter 3, corresponding to a review article, presents the state of the art on fruit detection and sizing methods. Knowledge of these methos is required to address the objectives set out in Chapter 1. Chapters 4 and 5, each one corresponding to a software article, introduce the developed software tools as well as the implementations of the fruit size and weight prediction algorithms proposed in this thesis. Chapter 6, corresponding to a research article, presents the experimental field evaluation of these algorithms. Chapter 7 discusses the results and establishes the lines of future work. Finally, general conclusions of this research are summarized in Chapter 8.

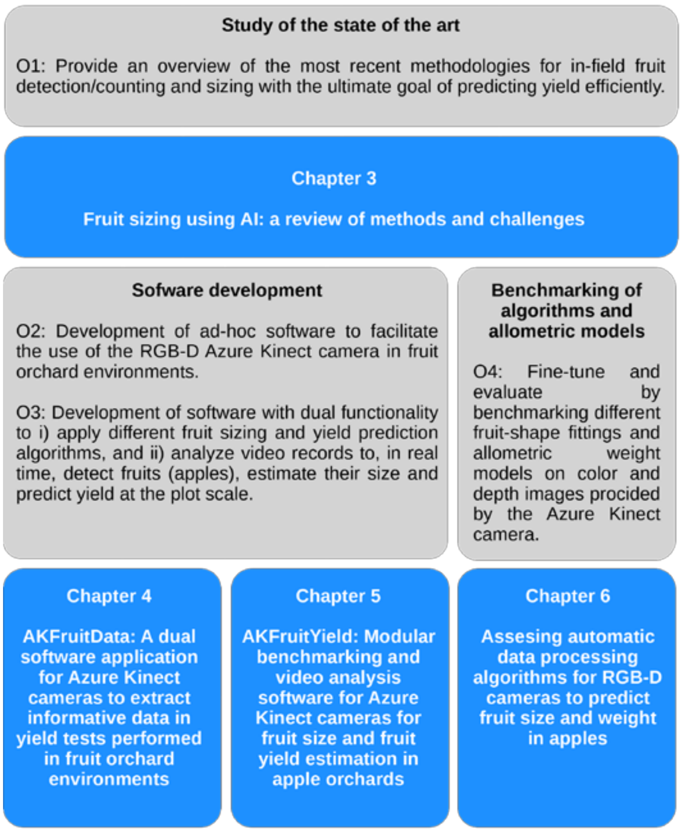


Fig. 1.1 Structure of the thesis with different chapters, objectives and articles published in journals.

1.6 References

Avendano, J., Ramos, P.J., Prieto, F.A., 2017. A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device. Expert Syst. Appl. 88, 178–192. https://doi.org/10.1016/j.eswa.2017.06.044

BOE, 2005. Spanish Ministerial Order PRE/1950/2005 [WWW Document]. URL https://www.boe.es/eli/es/o/2005/06/17/pre1950/con (accessed 7.7.21).

Chaudhuri, S., Kotwal, K., 2014. Hyperspectral image fusion, 1st ed, Hyperspectral Image Fusion. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-7470-8

Corti, A., Giancola, S., Mainetti, G., Sala, R., 2016. A metrological characterization of the Kinect V2 time-of-flight camera. Rob. Auton. Syst. 75, 584–594. https://doi.org/10.1016/j.robot.2015.09.024

Dong, W., Roy, P., Isler, V., 2018. Semantic Mapping for Orchard Environments by Merging Two-Sides Reconstructions of Tree Rows. J. F. Robot. 37, 97–121. https://doi.org/10.1002/rob.21876

Dorj, U.-O., Lee, M., Yun, S., 2017. An yield estimation in citrus orchards via fruit detection and counting using image processing. Comput. Electron. Agric. 140, 103–112. https://doi.org/10.1016/j.compag.2017.05.019

Eizentals, P., Oka, K., 2016. 3D pose estimation of green pepper fruit for automated harvesting. Comput. Electron. Agric. 128, 127–140. https://doi.org/10.1016/j.compag.2016.08.024

Feng, J., Zeng, L., He, L., 2019. Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. Sensors 19, 949. https://doi.org/10.3390/s19040949

Feng, Q., Chen, J., Li, X., Li, C., Wang, X., 2019. Multi-spectral image fusion method for identifying similar-colored tomato organs, in: 2019 IEEE International Conference on Unmanned Systems and Artificial Intelligence (ICUSAI). IEEE, pp. 142–145. https://doi.org/10.1109/ICUSAI47366.2019.9124912

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

Furukawa, Y., Hernández, C., 2015. Multi-View Stereo: A tutorial. Foundations and trends. Now Publ. Inc. 9, 1–148. https://doi.org/10.1561/9781601988379

Gan, H., Lee, W.S., Alchanatis, V., Abd-Elrahman, A., 2020. Active thermal imaging for immature citrus fruit detection. Biosyst. Eng. 198, 291–303. https://doi.org/10.1016/j.biosystemseng.2020.08.015

Gan, H., Lee, W.S., Alchanatis, V., Ehsani, R., Schueller, J.K., 2018. Immature green citrus fruit detection using color and thermal images. Comput. Electron. Agric. 152, 117–125. https://doi.org/10.1016/j.compag.2018.07.011

Gené-Mola, J., Gregorio, E., Guevara, J., Auat, F., Sanz-Cortiella, R., Escolà, A., Llorens, J., Morros, J.R., Ruiz-Hidalgo, J., Vilaplana, V., Rosell-Polo, J.R., 2019a. Fruit detection in an apple orchard using a mobile terrestrial laser scanner. Biosyst. Eng. 187, 171–184. https://doi.org/10.1016/j.biosystemseng.2019.08.017

Gené-Mola, J., Llorens, J., Rosell-Polo, J.R., Gregorio, E., Arnó, J., Solanelles, F., Martínez-Casasnovas, J.A., Escolà, A., 2020. 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., Vilaplana, V., Rosell-Polo, J.R., Morros, J.R., Ruiz-Hidalgo, J., Gregorio, E., 2019b. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. Comput. Electron. Agric. 162, 689–698. https://doi.org/10.1016/j.compag.2019.05.016

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

Häni, N., Roy, P., Isler, V., 2020. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. J. F. Robot. 37, 263–282. https://doi.org/10.1002/rob.21902

Khanal, S., Kushal, K.C., Fulton, J.P., Shearer, S., Ozkan, E., 2020. Remote sensing in agriculture—accomplishments, limitations, and opportunities. Remote Sens. 12, 1–29. https://doi.org/10.3390/rs12223783

Klette, R., 2014. Concise Computer Vision, 1st ed, Undergraduate Topics in Computer Science. Springer London, London. https://doi.org/10.1007/978-1-4471-6320-6

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, 2031–2038. https://doi.org/10.1109/LRA.2020.2970654

Li, L., Zhang, Q., Huang, D., 2014. A review of imaging techniques for plant phenotyping. Sensors (Switzerland) 14, 20078–20111. https://doi.org/10.3390/s141120078

Lin, G., Tang, Y., Zou, X., Xiong, J., Fang, Y., 2020. Color-, depth-, and shape-based 3D fruit detection. Precis. Agric. 21, 1–17. https://doi.org/10.1007/s11119-019-09654-w

Liu, G., Mao, S., Kim, J.H., 2019. A mature-tomato detection algorithm using machine learning and color analysis. Sensors (Switzerland) 19, 1–19. https://doi.org/10.3390/s19092023

Liu, X., Chen, S.W., Aditya, S., Sivakumar, N., Dcunha, S., Qu, C., Taylor, C.J., Das, J., Kumar, V., 2018. Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 1045–1052. https://doi.org/10.1109/IROS.2018.8594239

Liu, X., Chen, S.W., Liu, C., Shivakumar, S.S., Das, J., Taylor, C.J., Underwood, J., Kumar, V., 2019. Monocular Camera Based Fruit Counting and Mapping with Semantic Data Association. IEEE Robot. Autom. Lett. 4, 2296–2303. https://doi.org/10.1109/LRA.2019.2901987

Lu, B., Dao, P.D., Liu, J., He, Y., Shang, J., 2020. Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens. 12, 2659. https://doi.org/10.3390/RS12162659

Martinez-Guanter, J., Ribeiro, Á., Peteinatos, G.G., Pérez-Ruiz, M., Gerhards, R., Bengochea-Guevara, J.M., Machleb, J., Andújar, D., 2019. Low-Cost Three-Dimensional Modeling of Crop Plants. Sensors 19, 2883. https://doi.org/10.3390/s19132883

Méndez, V., Pérez-Romero, A., Sola-Guirado, R., Miranda-Fuentes, A., Manzano-Agugliaro, F., Zapata-Sierra, A., Rodríguez-Lizana, A., 2019. In-field estimation of orange number and size by 3D laser scanning. Agronomy 9. https://doi.org/10.3390/agronomy9120885

Milella, A., Marani, R., Petitti, A., Reina, G., 2019. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. Comput. Electron. Agric. 156, 293–306. https://doi.org/10.1016/j.compag.2018.11.026

Narvaez, F.Y., Reina, G., Torres-Torriti, M., Kantor, G., Cheein, F.A., 2017. A survey of ranging and imaging techniques for precision agriculture phenotyping. IEEE/ASME Trans. Mechatronics 22, 2428–2439. https://doi.org/10.1109/TMECH.2017.2760866

Onishi, Y., Yoshida, T., Kurita, H., Fukao, T., Arihara, H., Iwai, A., 2019. An automated fruit harvesting robot by using deep learning. ROBOMECH J. 6, 2–9. https://doi.org/10.1186/s40648-019-0141-2

Özyeşil, O., Voroninski, V., Basri, R., Singer, A., 2017. A survey of structure from motion. Acta Numer. 26, 305–364. https://doi.org/10.1017/S096249291700006X

Pajares, G., García-Santillán, I., Campos, Y., Montalvo, M., Guerrero, J.M., Emmi, L., Romeo, J., Guijarro, M., Gonzalezde-Santos, P., 2016. Machine-vision systems selection for agricultural vehicles: A guide. J. Imaging 2. https://doi.org/10.3390/jimaging2040034

Qureshi, W.S., Payne, A., Walsh, K.B., Linker, R., Cohen, O., Dailey, M.N., 2017. Machine vision for counting fruit on mango tree canopies. Precis. Agric. 18, 224–244. https://doi.org/10.1007/s11119-016-9458-5

Rose, J.C., Paulus, S., Kuhlmann, H., 2015. Accuracy analysis of a multi-view stereo approach for phenotyping of tomato plants at the organ level. Sensors (Switzerland) 15, 9651–9665. https://doi.org/10.3390/s150509651

Rosell, J.R., Sanz, R., 2012. A review of methods and applications of the geometric characterization of tree crops in agricultural activities. Comput. Electron. Agric. 81, 124–141. https://doi.org/10.1016/j.compag.2011.09.007

Sarbolandi, H., Lefloch, D., Kolb, A., 2015. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. Comput. Vis. Image Underst. 139, 1–20. https://doi.org/10.1016/j.cviu.2015.05.006

Sarkar, M., Theuwissen, A., 2013. Design of a CMOS Image Sensor, in: A Biologically Inspired CMOS Image Sensor. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 49–104. https://doi.org/10.1007/978-3-642-34901-0_3

Si, Y., Liu, G., Feng, J., 2015. Location of apples in trees using stereoscopic vision. Comput. Electron. Agric. 112, 68–74. https://doi.org/10.1016/j.compag.2015.01.010

Solomon, C., Breckon, T., 2011. Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab, 1st ed, Fundamentals of Digital Image Processing: A Practical Approach with

Examples in Matlab. Wiley Publishing. https://doi.org/10.1002/9780470689776

Stein, M., Bargoti, S., Underwood, J., 2016. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. Sensors (Switzerland) 16. https://doi.org/10.3390/s16111915

Szeliski, R., 2011. Computer Vision, 1st ed, Texts in Computer Science. Springer London, London. https://doi.org/10.1007/978-1-84882-935-0

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, 1–18. https://doi.org/10.3390/RS12152481

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H., Xue, Y., 2020. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. Precis. Agric. 21, 1072–1091. https://doi.org/10.1007/s11119-020-09709-3

Underwood, J.P., Hung, C., Whelan, B., Sukkarieh, S., 2016. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. Comput. Electron. Agric. 130, 83–96. https://doi.org/10.1016/j.compag.2016.09.014

Vit, A., Shani, G., 2018. Comparing RGB-D sensors for close range outdoor agricultural phenotyping. Sensors (Switzerland) 18. https://doi.org/10.3390/s18124413

Vitzrabin, E., Edan, Y., 2016. Changing Task Objectives for Improved Sweet Pepper Detection for Robotic Harvesting. IEEE Robot. Autom. Lett. 1, 578–584. https://doi.org/10.1109/LRA.2016.2523553

Williams, H.A.M., Jones, M.H., Nejati, M., Seabright, M.J., Bell, J., Penhall, N.D., Barnett, J.J., Duke, M.D., Scarfe, A.J., Ahn, H.S., Lim, J., MacDonald, B.A., 2019. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. Biosyst. Eng. 181, 140–156. https://doi.org/10.1016/j.biosystemseng.2019.03.007

Zhang, B., Huang, W., Wang, C., Gong, L., Zhao, C., Liu, C., Huang, D., 2015. Computer vision recognition of stem and calyx in apples using near-infrared linear-array structured light and 3D reconstruction. Biosyst. Eng. 139, 25–34. https://doi.org/10.1016/j.biosystemseng.2015.07.011

Zhang, C., Zhang, K., Ge, L., Zou, K., Wang, S., Zhang, J., Li, W., 2021. A method for organs classification and fruit counting on pomegranate trees based on multi-features fusion and support vector machine by

3D point cloud. Sci. Hortic. (Amsterdam). 278, 109791. https://doi.org/10.1016/j.scienta.2020.109791

Zhang, Q., Liu, Y., Gong, C., Chen, Y., Yu, H., 2020. Applications of deep learning for dense scenes analysis in agriculture: A review. Sensors (Switzerland) 20, 1520. https://doi.org/10.3390/s20051520

Zhao, Y., Gong, L., Huang, Y., Liu, C., 2016. A review of key techniques of vision-based control for harvesting robot. Comput. Electron. Agric. 127, 311–323. https://doi.org/10.1016/j.compag.2016.06.022

Chapter 2: Methodology

2.1 *Azure Kinect DK RGB-D Camera Basics*

A brief introduction to RGB-D sensors was made in Chapter 1, including some ideas about their advantages and drawbacks compared to other types of passive sensors. In particular, the Azure Kinect camera (Fig. 2.1) was introduced as an important component of the RGB-D family sensors, having also highlighted scientific works that have addressed and positively valued the use of this camera(Neupane et al., 2021; Pasinetti et al., 2023). Generally speaking, the Azure Kinect camera is perceived as a good digital option for fruit detection tasks in fruit orchard environments, although further research is still needed to address data acquisition and processing from depth cameras in fruit growing.

Fig. 2.1 a) Azure Kinect camera. b) Camera and accessories package. c) Camera mounted on the scene.

This PhD thesis exclusively addresses the use of the Azure Kinect camera in agriculture, specifically, for the purpose of detecting fruits and efficiently estimating yield. The Azure Kinect combines a 1-megapixel time-of-flight (ToF) camera, a CMOS rolling shutter sensor, an inertial measurement unit (IMU) and a microphone array. Regarding configuration of use in field conditions, experiments were carried out by collecting RGB, IR and depth data. So, both the operation of the IMU unit and the microphone remained disabled throughout our data collection, having set the depth camera mode to the narrow field-of-view (NFOV) unbinned option. Additional basic Azure Kinect specifications are shown in Table 2.1 (Microsoft, 2022).

Table 2.1. 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

2.2 *Software design bases for the Azure Kinect camera for use in fruit detection, sizing and yield prediction*

Acquiring data and extracting analyzable information is the first challenge that must be faced when using a depth camera in field conditions. Apart from knowing the state of the art of current technologies in fruit production (a review task that has given rise to Chapter 3), this PhD thesis fundamentally first addresses the design of several open source tools to acquire data and extract analyzable images using the Azure Kinect camera. In a second block of tools, the objective was to apply trained automatic detectors, sizing algorithms and allometric models for final weight prediction (yield) in apple trees. Additionally, uncertainty analysis of the entire process was also an important goal in order to assess the progress and reliability of the research.

A computer vision system with versatile use for different sensors (cameras) was the initial idea to achieve. This meant facing RGB-D cameras from different manufacturers with different technologies (ToF, stereo vision), and the additional requirement of importing georeferenced locations of detected fruits through a GNSS receiver. Both a static or mobile platform located in front of the tree canopies allowed the set of devices for experimentation to be housed. In the case of the mobile platform, the Azure Kinect camera was moved along the alley-ways of the tree plantation taking video records from the side of the apple tree canopies. Having collected data from the orchard, dataset creation was the next stage to address. For this, a new software module was programmed to allow data extraction functionalities (frames with previously labeled apples), thus constituting a separate software from the previous one that is used exclusively for video data acquisition. As can be seen later, Chapter 4 shows in greater detail the design, structure and functionalities of these tools.

Data was therefore made available. At this point, the second modular software had the purpose of training automatic fruit detectors and fruit sizing algorithms among other functionalities. A first module was conceived with the exclusive objective of benchmarking of fruit sizing and yield prediction algorithms. Color and depth images were used as inputs, allowing different sizing algorithms and allometric models to be combined to estimate the weight of apples. The second software module worked directly with video records, allowing different deep learning-based automatic fruit detectors to be implemented. By adding the optimal combination of sizing and yield allometric modeling (Chapter 6), plot-scale yield prediction should be possible. Previous design of these tools is shown in Chapter 5.

2.3 *Programming environment*

Once both the camera and the software functionalities were established, attention was focused on the programming language and environment. The Azure Kinect camera required a host computer for operational functionality under Windows or Ubuntu Linux operating systems, in addition to management drivers and the Software Development Kit (SDK). Choosing the programming language was a key point for the development of the software used in this thesis. C/C++ language was the default option facing the SDK programming. However, it was thought that a cross-platform programming language (adaptable to multiple operating systems) was a better option, opening up the possibility of integrating the use of Azure Kinect camera data (image data) with artificial intelligence algorithms.

In addition to the above, support from a broad development community to cover needs that may arise was also a quality to consider. For these reasons, the Python language was finally adopted, with multiple and valued advantages in syntax, support for libraries (image processing, deep learning, statistics) and online documentation. Certainly, the use of Python raised the need to select third-party libraries for acquiring and extracting data from the Azure Kinect camera. But, this issue was timely resolved by using Pyk4a (Asselin et al., 2021).

Figure 2.2 shows the flow chart of the development of the software tools and the milestones reached with each of them. AKFruitData_1 intended to be the tool focused on data acquisition using the Azure Kinect camera. Specifically, fruit tree canopies were recorded laterally providing videos for later processing. In a next step, AKFruitData_2 was intended to solve the problem of extracting frames and make available the necessary data to later apply fruit (apples) sizing algorithms. The second software tool package (AKFruitYield, Chapter 5) was conceived as a complement to the functionalities of AKFruitData (Chapter 4). Starting from the tree canopy images and having manually labeled apples within the frames, the AKFruitYield_1 tool was proposed to benchmark different sizing algorithms to analyze the best strategy for non-occluded apples and occluded apples. Allometric models to predict apple weight were then proposed using the geometric sizing parameters as input variables. In fact, benchmarking of algorithms and models is the topic discussed in Chapter 6. Finally, AKFruitYield_2 was conceived as an informed results delivery software, allowing automatic apple detectors and optimized sizing and yield prediction algorithms to be implemented together and making it possible to use the Azure Kinect camera at the plot level. The latter software was intended as a tool to provide, nested with AKFruitData_1 software, real-time results based on videos recorded along the alley-ways. In short, this is the final purpose of the set of tools that have been developed in this Doctoral Thesis.

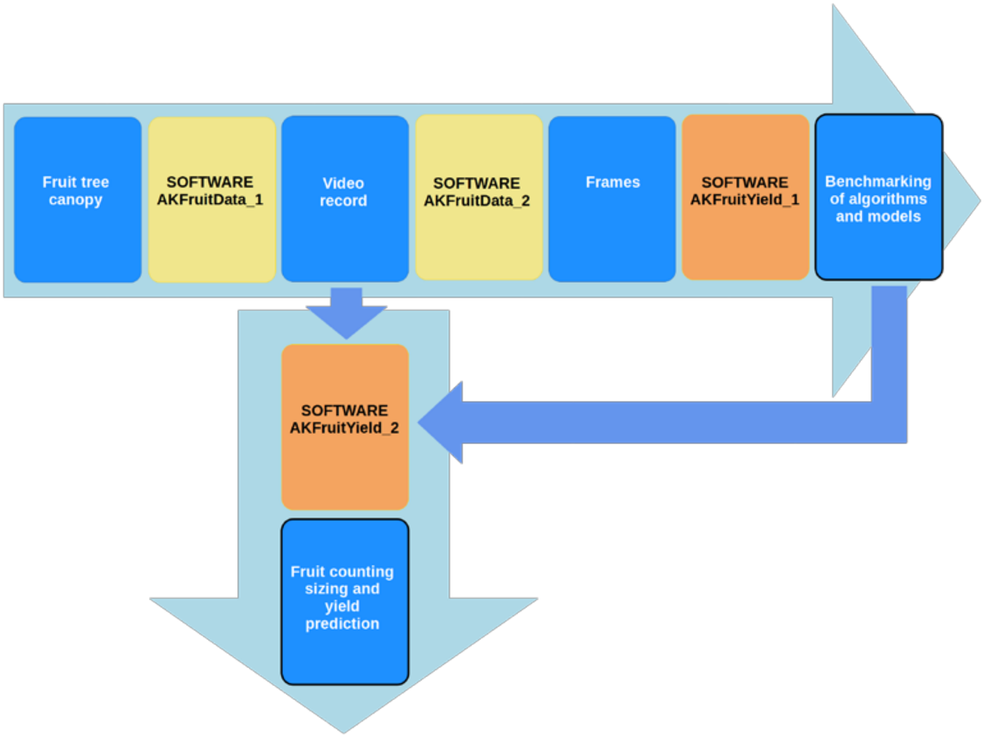


Fig. 2.2 Workflow of the proposed software solutions.

2.4 References

Asselin, L.-P., Martin, O., etiennedub, Gruzinov, I., DominikMa, Forstner, J. von, Boulanger, S., 2021. etiennedub/pyk4a: Version 1.2.4. https://doi.org/10.5281/ZENODO.4774311

Microsoft, 2022. Azure Kinect DK documentation [WWW Document]. URL https://docs.microsoft.com/en-us/azure/kinect-dk/ (accessed 9.19.23).

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

Pasinetti, S., Nuzzi, C., Luchetti, A., Zanetti, M., Lancini, M., De Cecco, M., 2023. Experimental Procedure for the Metrological Characterization of Time-of-Flight Cameras for Human Body 3D Measurements. Sensors 23, 538. https://doi.org/10.3390/s23010538

Chapter 3: Fruit sizing using AI: a review of methods and challenges

This chapter was published as:
Miranda, Juan C., Gené-Mola, J., Zude-Sasse, M., Tsoulias, N., Escolà, A., Arnó, J., Rosell-Polo, J.R., Sanz-Cortiella, R., Martínez-Casasnovas, J.A., Gregorio, E., 2023. Fruit sizing using AI: A review of methods and challenges. Postharvest Biol. Technol. 206, 112587. https://doi.org/10.1016/j.postharvbio.2023.112587

Fruit sizing using AI: a review of methods and challenges

Juan C. Miranda[1], Jordi Gené-Mola[1], Manuela Zude-Sasse[2], Nikos Tsoulias[2], Alexandre Escolà[1], Jaume Arnó[1], Joan R. Rosell-Polo[1], Ricardo Sanz-Cortiella[1], José A. Martínez-Casasnovas1, Eduard Gregorio[1]*

[1]Research Group in AgroICT & Precision Agriculture, Universitat de Lleida (UdL) - Agrotecnio CERCA Center, 25198 Lleida, Catalonia, Spain

[2]Department of Horticultural Engineering, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee, 14469 Potsdam, Germany

*Corresponding author. E-mail address: eduard.gregorio@udl.cat (E. Gregorio).

Abstract

Fruit size at harvest is an economically important variable for high-quality table fruit production in orchards and vineyards. In addition, knowing the number and size of the fruit on the tree is essential in the framework of precise production, harvest, and postharvest management. A prerequisite for analysis of fruit in a real-world environment is the detection and segmentation from background signal. In the last five years, deep learning convolutional neural network have become the standard method for automatic fruit detection, achieving F1-scores higher than 90 %, as well as real-time processing speeds. At the same time, different methods have been developed for, mainly, fruit size and, more rarely, fruit maturity estimation from 2D images and 3D point clouds. These sizing methods are focused on a few species like grape, apple, citrus, and mango, resulting in mean absolute error values of less than 4 mm in apple fruit. This review provides an overview of the most recent methodologies developed for in-field fruit detection/counting and sizing as well as few upcoming examples of maturity estimation. Challenges, such as sensor fusion, highly varying lighting conditions, occlusions in the canopy, shortage of public fruit datasets, and opportunities for research transfer, are discussed.

Keywords: artificial intelligence, fruit detection, fruit measure, image processing, deep learning, fruit quality

3.1 *Introduction*

Agricultural production of fresh fruit and vegetables must substantially increase to address the food demand due to the growing population, which is expected to reach 9.7 billion people by mid-century (UN, 2022). However, the production needs to respect the environment and meet the requirements of social and economic sustainability (FAO, 2017). Avoiding food waste is a major concern in all these aspects. Food waste in the fruit supply chain can be caused by lack of fruit safety and decay of fruit, as well as rejection of fruit on the market due to insufficient product quality (Nicastro and Carillo, 2021). Product decay is addressed by postharvest technologies to keep fruit at marketing quality. On the other hand, achieving the desired fruit quality is left to the farmers, but demanded by the actors of the value chain (Saitone and Sexton, 2017). If the market value is considered, the appearance represents the most important quality parameter that needs to be achieved (Musacchi and Serra, 2018). The main variable of complex appearance is the fruit size. Size as well as other variables of appearance such as colour, shape or absence of defects are addressed by means of inline grading, sorting fruit according to the different market needs. However, too large fruit are difficult to market, since they frequently show reduced storability (Paul and Pandey, 2014), whereas consumers prefer large, but not uncommonly large fruit (Iwanami, 2011). Similarly, it is difficult to find an economically reasonable market for small fruit, due to unfavourable ratio of edible to residual parts of the fruit.

Size is considered a “search characteristic” that can be assessed before purchasing (Yeo and Edwards, 2006). It plays a role in the consumer’s decision to buy. Consumers preference for apples varies depending on country, region, type of market, gender, family income, education, age, food safety factors (i.e. pesticide use) and memory of previous eating experiences (Harker et al., 2003; Bonany et al., 2013; Bavay et al., 2013; Favre et al., 2022). A study conducted in Canadian territories reported the ideal size for dessert apple ranges between 74 and 76 mm considering various ages (Hampson et al., 2002). Sorting machines in packing lines are commonly dividing fruit in size classes and the class obtained influences the fruit price. Apple size is frequently included in trading standards or regional producers’ standards. In the European market, cultivar-specific fruit size of 60 mm has been requested (OECD, 2021). However, such requests have been becoming less binding due to introduction of kids’ apples and other specific products. Nevertheless, for sweet cherry, but also European plum and other stone fruits, the size of fruit is directly affecting the market price, e.g. in sweet cherry cultivar 'Celeste' even selective harvesting appears as economically vital. It was shown that at red ripening stage 55 days after full bloom (DAFB), 35 % of cherries reached >28 mm, whereas after 60 DAFB another 40 % more passed this value-creating threshold. Selective harvesting was reasonable in this case, since the higher size class gained 20 % increased market price (Heim and Zude-Sasse, 2014; He et al., 2015). However, for drawing harvest conclusions, the information on fruit size is requested in real-time.

Strategies to produce the desired fruit size are captured in the concept of crop load management (Robinson et al., 2017), which requires the feedback on actual fruit number and fruit size in the field (Delong et al., 2004). High crop load, beyond the fruit bearing capacity may result in small fruit considering the MaluSim approach from Lakso (Penzel et al., 2020). Low crop load situation obviously results in reduced yield per area, however, yield can be even further reduced due to an increased risk of storage disorders appearing in large fruit. In control and 1-MCP treated 'Gala' apples flesh breakdown in storage increased with enhanced fruit size (Lee et al., 2013). In nectarine, fruit size served as an input variable to model fruit development and storability (Casagrande et al., 2021). Furthermore, in non-destructive quality sensing of citrus fruit, it was shown that fruit size affects the non-destructive spectral-optical analysis in the short-wave near infrared (NIR) wavelength range (Miller and Zude-Sasse, 2004; Sun et al., 2021). Fraser et al. (2003) has shown that the distribution of light varies within citrus fruit. Consequently, information on fruit size and fruit size distribution in the canopy may support development of more robust sensor calibrations.

Automatic detection, location and sizing of fruit in the field are agricultural problems in which computer vision and geo-positioning play a fundamental role. Fruit detection consists of finding a candidate region of interest (ROI) in a given image, point cloud or other type of data, and classifying it as fruit or background. The fruit location problem goes even one step further by locating the fruit in a local or global coordinate system (e.g. the position of the fruit in an image or on the Earth, respectively), and making a coordinate conversion to transfer fruit position detected in images onto the real world coordinate system. In-field fruit sizing consists of measuring the fruit (e.g. diameter, length, volume, etc.) on the tree to obtain morphological data.

Systems applied to fruit detection and sizing must deal with data acquired under a variety of lighting conditions (Chaivivatrakul and Dailey, 2014), and their performance may be affected by factors such as shadows, reflections, backlights, background colour, inclusion and occlusions (Fig. 3.1a,b). Other factors such as coinciding structures or the slope in an orchard (Fig. 3.1c) affect the open view to the fruit object and require geometric correction of sensor raw data. The combination of these factors influences the accuracy of the detection result. According to the lighting conditions, it is useful to mention work carried out in night-time conditions (Fig. 3.1d), with the help of artificial lighting rigs used both to lighten the scene and to reduce the undesirable effects of variable lighting. Fruit clustering, occlusions and shading (Fig. 3.1e,f) are other factors that need to be taken into consideration in fruit detection (Jarvinen et al., 2019).

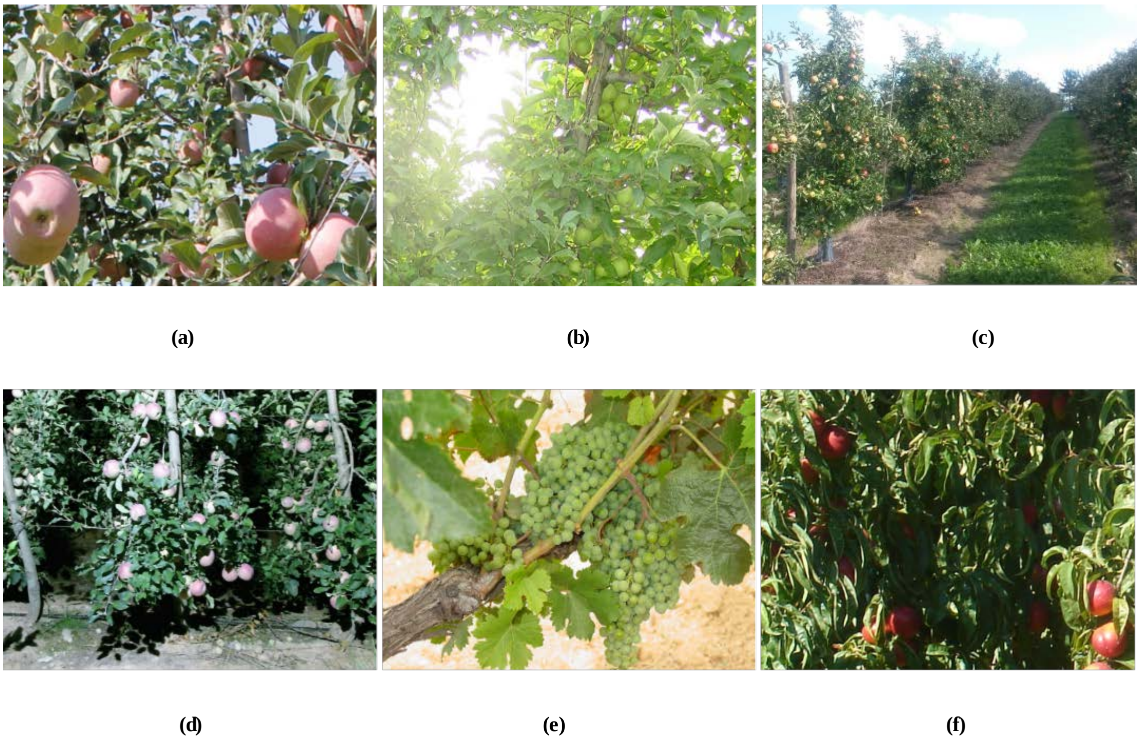


Fig. 3.1. Examples of fruit detection challenges under several conditions. (a) Fuji apples with shadows (Gené-Mola et al., 2020e). (b) Golden Delicious apples (colour similar to the background) with reflections and backlights. (c) Slope in an apple tree orchard. (d) Apple tree in night-time conditions (Gené-Mola et al., 2019b). (e) Cluster of grapes (Arnó, 2008). (f) Peaches, occlusions and shading.

Regarding fruit growing stages, changes in colour and shape along the growing season affect the performance of the fruit detection system. Depending on the purpose of the study, measurements are taken at different stages: mapping of flower and fruit distributions for yield prediction (Underwood et al., 2016), or detection of fruit at several stages of growth to monitor the evolution of the orchard (Tian et al., 2019; Tsoulias et al., 2022). Nevertheless, most studies have been carried out at harvest time, when the objective is, for example, to predict the yield, map the production or for automatic harvesting purposes (Gené-Mola et al., 2019a; Wang et al., 2019). In this stage, the fruit has reached maximum size and frequently changed colour from green to yellow or red, which is less challenging compared to detection of small, green objects in green foliage (Tsoulias et al., 2020, 2023).

With respect to the algorithms used, two aspects need to be considered: 1) obtaining a high-performance fruit detection, which means having high detection rates and a low number of false positives; 2) the development of computationally efficient algorithms to achieve low processing times (Häni et al., 2020a). In this regard, the application of deep learning in computer vision and the use of 3D sensors have revolutionized fruit detection (Koirala et al., 2019a). However, the shortage of public fruit datasets, as well as the diversity of lighting conditions and capture devices, makes it difficult to compare the fruit detection algorithms that have been published (Qureshi et al., 2017). Nonetheless, efforts have been carried out to collect and classify specialized agricultural datasets that include different sensor types, fruit varieties, and field conditions (Lu and Young, 2020). In addition, it should be noted that fruit detection and sizing systems usually deal with complex, unstructured and changing agricultural environments, in contrast to the generally clearly defined targets that detection systems work on in industrial applications (Bechar and Vigneault, 2016; Zhao et al., 2016a). Although promising results have been achieved in industry environments, it is still cumbersome to determine the fruit load when these techniques are implemented in the field. All the reasons set out above explain why fruit detection and sizing appears an interesting application of AI and is currently a focal point of interest.

This work presents a review of the state-of-the-art of computer vision-based fruit detection and sizing methods. The present review work is structured in six main sections. Section 3.1 comprises this introduction. Sections 3.2 and 3.3 deal with fruit detection, reviewing the handcrafted computer vision and deep learning methods, respectively. Section 3.4 covers the field of fruit size analysis and maturity estimation that can be later applied in crop load management and yield estimation. In section 3.5, the challenges to be faced when applying fruit sizing are discussed. Final conclusions are presented in section 3.6.

3.2 Fruit detection based on handcrafted features

3.2.1 Background

Before the advent of deep learning, most of the computer vision algorithms relied on the identification and extraction of image features such as corners, edges and blobs, and the subsequent classification of these features that defined the image or parts of the image. The design of the methodology to extract these features was done manually (handcrafted) based on human vision insights and intuitions (Nanni et al., 2017); this is why these algorithms are known as handcrafted feature-based methods. Previous reviews of these methods are thoroughly described in Gongal et al. (2015) and in Zhao et al. (2016a). While there is no single recipe to frame all the handcrafted methods, the aim of this section is to provide an up-to-date review.

Fruit detection algorithms, as a special case of general object detection (Fig. 3.2), can follow approaches based on two main steps (Wang and Zheng, 2019; Ward et al., 2019): (1) candidate region proposals generation; and (2) detection and recognition.

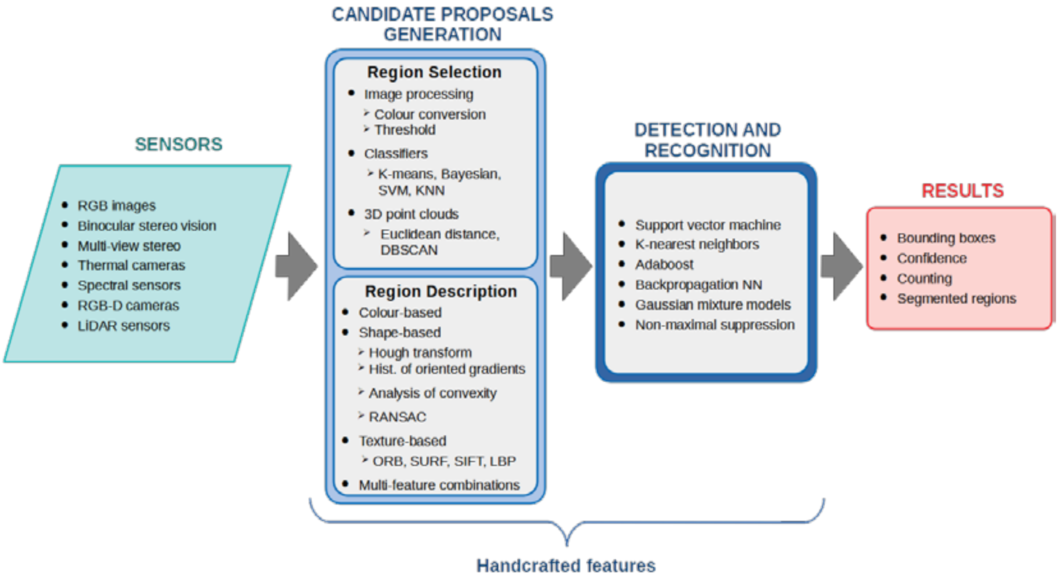


Fig. 3.2. Pipeline for fruit detection based on handcrafted features.

3.2.2 Candidate region proposals

The generation of candidate region proposals is the step of the process in which potential regions of interest are identified from data received by sensors. As sub-tasks, this can be divided into region selection and region description. Thresholding has been one of the most commonly used methods to classify fruit and background regions (Fig. 3.3). This method aims to binarize data by setting a numerical threshold into a discriminative feature that describes the object of interest. A common feature used has been the colour (Maldonado and Barbosa, 2016; Qian et al., 2018), although other types of data have been considered such as the area of pixels (Liu et al., 2019), the depth (Tao and Zhou, 2017), and the temperature using thermal imaging (Pedraza et al., 2019). Fruit reflectance and geometric features are also applied to define the fruit ROI (Gene-Mola et al., 2019a; Tsoulias et al., 2020).

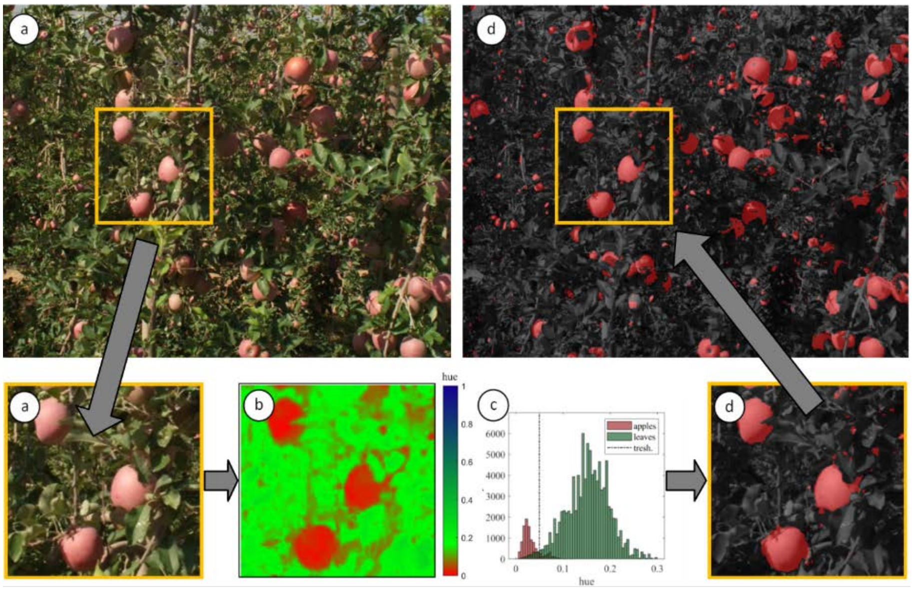


Fig. 3.3. Example of colour conversion and intensity thresholding applied to fruit segmentation. (a) RGB image. (b) Image converted to the HSV colour space. Colour scale corresponds to the hue (H) value. (c) Histogram of hue values for apple (red) and background (green) pixels. The vertical dash-dotted line corresponds to the selected threshold. (d) Segmented apples after applying the hue threshold.

Another possibility for region selection is to apply machine learning classifiers. Classification methods allow objects within a space to be distinguished by specific features. The most used classifiers include the unsupervised k-means algorithm (Wang et al., 2018a; Shi et al., 2020), and different supervised methods such as Bayesian (Lin et al., 2019), the k-nearest neighbours (KNN) (Qureshi et al., 2017) and support vector machine (SVM) procedures (Zhang et al., 2020). With regard to 3D point clouds, there are many methods that allow their segmentation (Grilli et al., 2017). Two of the most commonly used methods in fruit detection are Euclidean clustering (Nguyen et al., 2016) and density-based spatial clustering of applications with noise (DBSCAN) (Eizentals and Oka, 2016).

Region description is a step prior to detection and recognition in which the identified regions are described with features to refine the selection according to their appearance and geometry. The result can be a multi-dimensional numeric vector or a set of pixels or point cloud with candidate labels (fruit, background, etc.). Colour, shape, texture and multiple features are used to describe regions.

Colour-based radiometric features mostly comprise the statistical data about channels in colour spaces. For example, in Syal et al. (2014) features were extracted by using the mean colour of the ‘a’ and ‘b’ components in L*a*b space. Using light detection and ranging (LiDAR) at 660 nm provides information on the chlorophyll content of fruit, which can support the segmentation (Tsoulias et al., 2023).

Shape-based techniques are useful in cases where the fruit and the background have the same colour. These are ideal for detecting fruits whose shape differs from leaves and branches. In fruit detection, the most relevant shape-based techniques are the Hough transform (HT) and the histogram of oriented gradients (HOG). One of the main variants of HT is the circular Hough transform (CHT), which has been widely used to locate spherical fruit in orchards (Wang et al., 2018a; Chen et al., 2021). Other shape-based techniques include analysis of convexity (Kelman and Linker, 2014), three-point circle fitting (Sun et al., 2019), or random sample consensus (RANSAC) (Nguyen et al., 2016).

Textures are small patterns with fluctuations of the intensity between groups of neighbouring pixels. Texture-based methods are used to detect fruit of the same colour as the background, taking advantage of the invariant characteristics of textures to changes in lighting and the smoother surfaces of the fruits. Among the texture-based methods used in fruit detection can be cited oriented FAST and rotated BRIEF (ORB), speeded-up robust features (SURF), scale-invariant feature transform (SIFT) and local binary patterns (LBP) (Chaivivatrakul and Dailey, 2014; Wang et al., 2018a). Multiple feature combinations have been preferred by some authors to improve the detection success rate. Li et al. (2016) and Qureshi et al. (2017) present examples of this approach for the detection of immature citrus and mango fruit, respectively.

3.2.3 Detection and recognition

Once a set of candidate regions and a list of features that describe each of these regions have been obtained, the next step is to classify them into true (fruit) or false (background) detection. For this purpose, a variety of classifiers have been used such as SVM (Gené-Mola et al., 2020a; Wu et al., 2020), KNN (Li et al., 2016; Nyarko et al., 2018), Adaboost (Wang et al., 2018a; Mekhalfi et al., 2020), random forest (Yu et al., 2021), backpropagation neural network (BPNN) (Cheng et al., 2017), and Gaussian mixture model (GMM) (Roy et al., 2019).

A common source of error in fruit counting systems is the presence of multiple detections (more than one detection of a single fruit), which results in an over-counting error. To prevent multiple detections, some authors have applied the non-maximum suppression (NMS) algorithm, which consists of discarding the overlapped detections with non-maximum confidence values (Yu et al., 2021).

Handcrafted detection algorithms are still applied due to their lower use of resources (computer power and memory) (Zhang et al., 2020) and the relatively minor amount of data required to train them compared to blackbox methods such as deep neural networks. They are used in cases where the object has a high contrast with the background and can be easily distinguished. These methods have also been implemented on platforms where high computational power is not available (Fu et al., 2018; Habib et al., 2020). The disadvantage of handcrafted detection algorithms is the lack of generalization in detecting fruit in other acquisition conditions for which specific algorithms were not designed. In addition to this, the functions need to be optimized manually, which is time consuming (Farjon et al., 2020).

3.3 Fruit detection based on deep learning

3.3.1 Background

Deep learning has meant a breakthrough in computer vision and, consequently, in fruit detection. Koirala et al. (2019a) reviewed the use of deep neural networks for fruit detection. Prior to 19/01/2019 they found a total of 9 papers in the Scopus data base (www.scopus.com) using the keywords: 'deep' + 'learning' + 'fruit' + 'detection'. Four years later (on 31/07/2023), a total of 347 articles were found in Scopus on the same search basis, showing that the use of deep learning for fruit detection is a highly active research field with a rapid increase in scientific production (Fig. 3.4).

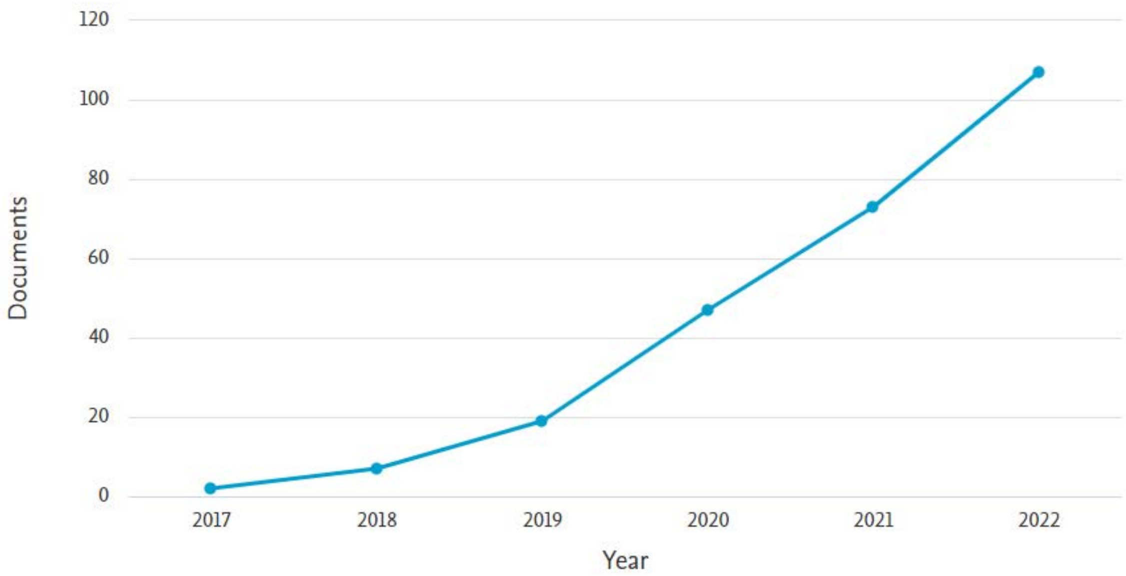


Fig. 3.4. Number of articles (conference proceedings not included) published per year in Scopus data-base containing keywords 'deep' + 'learning' + 'fruit' + 'detection'.

The most commonly used deep neural networks in computer vision are the so-called convolutional neural networks (CNN), where the neurons of each unit are organized in three-dimensional matrices (feature maps). Consecutive units are connected by means of convolutional layers, pooling layers and fully connected layers used to process the input data and extract features at different scales (LeCun et al., 2015).

CNNs have demonstrated a level of performance similar to that of the human eye in tasks such as image classification, object detection, and semantic and instance segmentation (Voulodimos et al., 2018). Image classification refers to the problem of classifying the whole image in a specific class, for instance an image of a fruit in the fruit class or variety (Fig. 3.5a). Object detection refers to the problem of identifying the regions (bounding boxes) that contain the objects of interest, for instance locating the fruit that appear in an image (Fig. 3.5b). Semantic segmentation refers to the problem of classifying each pixel in the image, for instance labelling each pixel as fruit, trunk, branch or background (Fig. 3.5c). Finally, instance segmentation combines object detection and semantic segmentation: first objects of interest are located in the image and then the objects are segmented, identifying which pixels of the image correspond to each detected object (Fig. 3.5d).

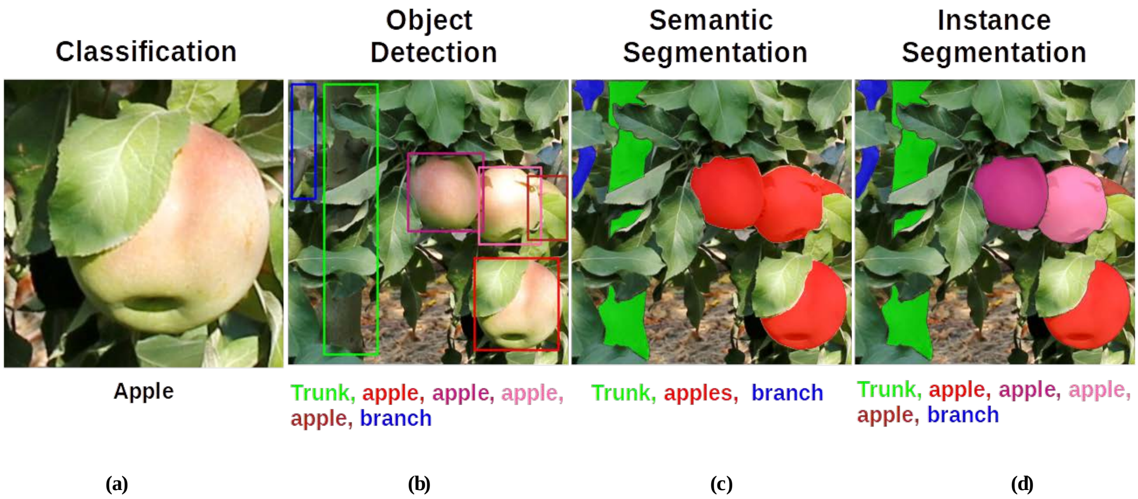


Fig. 3.5. Common computer vision tasks. Examples of apples on trees in field conditions: (a) classification, (b) object detection, (c) semantic segmentation, (d) instance segmentation.

A comparative table of the results reported in different deep learning-based fruit detection works is shown in Table 3.1. The F1-score metric was selected as it is the most commonly used in fruit detection papers. Other metrics such as average precision (AP) or accuracy (ACC) are reported when the F1-score results were not available. It should be noted that the reported results depend not only on the CNN structure and its parameters but also on the difficulty of the dataset. Thus, works assessing different structures with different datasets are not comparable.

Table 3.1. A comparative table of results reported in different deep learning-based fruit detection works. Results are reported in terms of F1-score and processing time per image. Accuracy (ACC), Pearson's R value and Average Precision (AP) are provided when the F1-score value is not available.

Approach	Data type	Method	Backbones	Crop	F1-score	Processing time (seconds per image)	Reference
Image classification	RGB	ResNet50	N/A	Apples	0.978 (ACC)	N/M	(Häni et al., 2020a)
	RGB	CountNet	VGG-16	Apples	0.962 (R)	N/M	(Bhattarai and Karkee, 2022)
Object detection	RGB	Faster-RCNN	VGG-16	Mangoes	0.881	N/M	(Stein et al., 2016)
	RGB	Faster-RCNN	VGG-16	Mangoes	0.908	0.13	(Bargoti and Underwood, 2017a)
	RGB	CNN + WS	Self-developed	Apples	0.861	0.24	(Bargoti and Underwood, 2017b)
	RGB	Faster-RCNN	ResNet-50	Strawberries	0.842	0.113	(Chen et al., 2019)
	RGB	MangoYOLO(pt)	N/M	Mangoes	0.968	0.015	(Koirala et al., 2019b)
	RGB	YOLOv2-M1	Darknet-19	Apples, pears	0.79	0.05	(Bresilla et al., 2019)
	RGB	YOLOv3dense	Darknet-53	Apples	0.864	0.304	(Tian et al., 2019)
	RGB	Faster-RCNN	Inception v2	Avocados	0.84 (AP)	0.217	(Vasconez et al., 2020)
	RGB	Faster-RCNN	Inception v2	Cherries	0.733	N/M	(Villacrés and Auat Cheein, 2020)
	RGB	Faster-RCNN	ResNet v2 Atrous	Apples	0.919	N/M	(Apolo-Apolo et al., 2020b)
	RGB	DY3TNet	Darknet-53	Kiwis	0.903 (AP)	0.034	(Fu et al., 2021)
	RGB	ATSS	ResNet50	Apples	0.925 (AP)	N/M	(Biffi et al., 2021)
	RGB	YOLOv4dense	DenseNet	Cherries	0.947	0.467	(Gai et al., 2021)
	RGB	YOLOv5s-pruned	Modified CSPDarknet	Apple fruitlets	0.915	0.008	(Wang and He, 2021)
	RGB	YOLOv5s-attention	Modified CSPDarknet	Apples	0.875	0.015	(Yan et al., 2021)
	RGB	YOLOv4	CSPDarknet-53	Bananas	0.941	0.045	(Fu et al., 2022)
	RGB+Thermal	Faster-RCNN + CHT	VGG-16	Oranges	0.929	N/M	(Gan et al., 2018)
	RGB+ NIR_C+Depth	Faster-RCNN	VGG-16	Apples	0.898	0.074	(Gené-Mola et al., 2019c)
	RGB+Depth	MS-FRCNN	ResNet101	Passion fruits	0.946	0.175	(Tu et al., 2020)
	RGB+Depth	NT-FFN	Self-developed	Citrus	0.934	0.026	(Sun et al., 2022)
Semantic segmentation	HyperSpectral	Hyperspectral CNN	N/A	Mangoes	0.989	N/M	(Wendel et al., 2018)
	RGB	FCN-8S	VGG-16	Kiwis	0.878	0.25	(Williams et al., 2019)
	RGB	MangoNet+CCL	N/A	Mangoes	0.844	N/M	(Kestur et al., 2019)
Fruit edge segmentation	RGB	Self-developed	ResNet 50	Apples	0.531	0.075	(Wang et al., 2020)
Instance segmentation	RGB	Mask-RCNN	ResNet50+FPN	Strawberries	0.956	0.125	(Yu et al., 2019)
	RGB	Mask-RCNN	ResNet101	Grapes	0.847	N/M	(Santos et al., 2020)
	RGB	Mask-RCNN	ResNet101-FPN	Apples	0.858	0.15	(Gené-Mola et al., 2020d)
	RGB	Mask-RCNN-suppression	ResNet101-FPN	Apples	0.905	0.25	(Chu et al., 2021)
	RGB	Mask-RCNN-attention	ResNet50+FPN	Apples	0.964	0.25	(Wang and He, 2022)
Multitask	RGB	DaSNet-v1	ResNet-101	Apples	0.832	0.072	(Kang and Chen, 2019)
	RGB	DaSNet-v2	Darknet-53	Apples	0.873	0.070	(Kang and Chen, 2020)
Point cloud segmentation	Point cloud	PointNet	PointNet	Grapes	0.91 (ACC)	N/A	(Kurtser et al., 2020a)
	Point cloud	LFPNet	PointNet	Apples, pears, grapes	0.802 (ACC)	N/A	(Yu et al., 2022a)
	Point cloud	Mask-RCNN	F-PointNet	Pomegranates	0.845	N/A	(Yu et al., 2022b)

N/A = not applicable. N/M = not mentioned. NIR_C = Near-infrared (range-corrected intensity).

3.3.2 Fruit detection using image classification CNNs

The structure of image classification CNNs is based on an input layer (the image to classify) connected with a group of convolutional layers that act as feature extractors (feature maps), ending with a group of fully connected layers that act as classifiers. The convolutional layers encode image features into more discriminative features by convolving the feature maps with filters (learned weights). Finally, fully connected layers are placed at the end of the CNN to classify feature maps in one of the classes of the output layer.

The use of classification CNNs for fruit counting is marginal because these architectures classify the entire image in a unique class and do not locate the objects inside images. Wang et al. (2021) proposed a modified version of the VGG16 network (Simonyan and Zisserman, 2014) to count the number of apple flowers in an image. The total number of flowers in the image was considered the image class, and the network was trained to directly estimate the number of flowers visible in the image, without locating them. A similar approach was used in Bhattarai and Karkee (2022), who modified the classification block of the VGG16 architecture to regress the number of flowers or fruits in apple tree images.

3.3.3 Fruit detection using object detection CNNs

Object detection CNNs are formed with two main structures: backbone and head. The backbone usually uses the first layers of an image classification CNN as feature extractor to encode the data into feature maps. Then, the head structure uses the feature maps provided by the backbone to predict the object locations and their class. Depending on the head structure, object detection networks can be classified as one- or two-stage networks.

The first CNNs used for fruit detection were two-stage networks type, with a structure based on two main modules: (1) a region proposal module used to propose ROIs likely to contain a fruit; (2) a classification branch used to classify the proposed regions into fruit or background and refine the detection bounding box. The most commonly used two-stage CNN for fruit detection is the Faster-RCNN (Ren et al., 2017), which has been used to detect apples (Apolo-Apolo et al., 2020b; Kang and Chen, 2020; Tian et al., 2019), oranges (Apolo-Apolo et al., 2020a, Biffi et al., 2021), mangos (Bargoti and Underwood, 2017a; Koirala et al., 2019b), kiwis (Gan et al., 2018), and strawberries (Chen et al., 2019), among others.

One-stage CNNs (or single shot detectors) simultaneously predict object class and bounding box without the need of a region proposal branch. Single shot detectors (SSD) used for fruit detection include the single shot multibox detector (Liu et al., 2016) and the You Only Look Once (YOLO) (Redmon and Farhadi, 2018) and its variants v2, v3, v4 and v5. The SSD was used in Vasconez et al. (2020) with the MobileNet backbone for detection of apples, avocadoes and lemons. YOLOv2, YOLOv3 and YOLOv4 were used with DarkNet-19 and DarkNet-53 backbones, respectively, in different fruit detection works for apples, pears, kiwis, mangoes, bananas and grapes (Bresilla et al., 2019; Fu et al., 2021, 2022; Koirala et al., 2019b; Santos et al., 2020; Tian et al., 2019).

To enhance the performance of fruit detection systems, some authors have proposed the use of multi-modal deep neural networks to fuse different image modalities such as colour (RGB), depth or infrared (IR) intensity. Using a red-green-blue-depth (RGB-D) camera, Gené-Mola et al. (2019c) showed an increase of 4.46 % in the F1-score when combining colour, range-corrected IR intensity and depth images for apple detection with Faster-RCNN. Similarly, colour and depth images were combined for passion fruit detection in Tu et al. (2020), and colour and thermal images were combined in Gan et al. (2018) for orange detection. More recently, Sun et al. (2022) developed a new multi-modal network termed noise-tolerant feature fusion network (NT-FFN) which merged colour and depth features by means of attention modules, resulting in a better fruit detection performance: from F1-score of 0.910 (using RGB) to 0.934 (fusing RGB-D through NT-FFN).

The introduction of edge computing applications and the need of deploying real-time fruit detection in embedded computers, such as NVIDA Jetson products, has shifted the attention of researchers to the development of smaller object detection CNNs (Roy and Bhaduri, 2022; Zhang et al., 2021b; Zhang et al., 2022a). In consequence, many fruit detection papers published during the last two years are focused on achieving faster inference speeds in low power devices by means of light-weight and fast CNNs such as YOLOv5s and other YOLO-based tiny variants (Gai et al., 2021; Wang and He, 2021; Yan et al., 2021).

3.3.4 Fruit detection using semantic and instance segmentation CNNs

The fully convolutional network (FCN) (Long et al., 2015) is one of the most used architectures for fruit segmentation. FCN uses the first convolutional layers of CNN image classification as a backbone to encode data in discriminative feature maps. Then, the last feature map from the backbone is up-scaled by means of skip connections that combine information from shallower layers (finer but less discriminative) and deeper layers (coarser but more discriminative). FCN was used to detect kiwi fruits (Williams et al., 2019), oranges and apples (Chen et al., 2017; Liu et al., 2018).

Other authors have opted to develop new architectures specifically designed for fruit segmentation. The MangoNet architecture, developed by Kestur et al. (2019), replaced the last 3 convolution layers of FCN with a single convolution layer, obtaining a similar performance to that of FCN but reducing network complexity. Wang et al. (2020) developed a new architecture to adapt ResNet-50 for apple edge segmentation. Bargoti and Underwood (2017b) proposed a sliding window approach, which classified each pixel by means of a self-developed multilayer perceptron (MLP) and a CNN.

A disadvantage of semantic segmentation CNNs is that it is not possible to directly count fruits from a segmented image because all fruits appearing in an image are segmented under the same class. Instance segmentation CNNs overcame this issue by combining object detection and semantic segmentation. The most popular instance segmentation CNN used for fruit detection is Mask-RCNN (He et al., 2017), which is an extension of Faster-RCNN that includes a segmentation branch to mask detected objects. Mask-RCNN was used with VGG-19 backbone for apple detection (Kang and Chen, 2020), with ResNet-50 backbone for strawberry detection (Yu et al., 2019) and with ResNet-101 backbone for apple (Gené-Mola et al., 2020d) and grape detection (Santos et al., 2020). More recently, some authors have proposed modifications in the Mask-RCNN architecture in order to achieve a better fruit detection performance (Chu et al., 2021; Wang and He, 2022) or a faster inference speed (Jia et al., 2021).

Kang and Chen (2019) developed a multi-task architecture termed "Detection and Segmentation Network" (DaSNetv1). This architecture combines a segmentation branch used to segment apples, trunks and branches, and a detection branch to locate fruits. This network was specifically designed for harvesting robots, allowing the detection of fruits and obstacles (branches) in a single network. Later, the same authors presented an improved version (DaSNetv2) (Kang and Chen, 2020) which replaced the previous detection branch with an instance segmentation architecture, allowing detection and instance segmentation to be performed on fruits, and semantic segmentation on branches in one step.

So far, the reviewed architectures were designed for working with image data. However, the evolution of photonics has led to the deployment of 3D sensors for robotic applications and, thus, to an increasing interest in using deep learning architectures to work with 3D data such as point clouds. Kurtser et al. (2020a) proposed the use of PointNet (Qi et al., 2017) for grape segmentation in 3D point clouds acquired with RGB-D sensors. The best results were obtained when combining RGB and XYZ data, reporting an average accuracy of 65 % in field conditions. Inspired by PointNet, Yu et al. (2022a) developed a new lightweight architecture for apple, pear and lemon point cloud segmentation that reported a mean accuracy of 80.2 %. Recently, Yu et al. (2022b) have tested the F-PointNet (Qi et al., 2018), a variant of the PointNet in which the frustrum between the camera shooting and the detected fruits is used for point cloud segmentation. An F1-score of 0.845 and AP score of 0.952 were obtained for mature pomegranate fruit detection.

3.3.5 Datasets for training fruit detection CNNs

The main disadvantage of using deep learning methods is the high amount of annotated data required for training models. The existence of large datasets such as ImageNet (Deng et al., 2009), Pascal VOC (Everingham et al., 2010) or COCO (Lin et al., 2014) enables CNN pre-training with publicly available data and fine-tuning of the network for fruit detection with new annotated images, reducing significantly the amount of images required to train the CNN. Nevertheless, the annotation of new data continues to be an intensive time-consuming task (Koirala et al., 2019b).

Some authors have analysed the correlation between dataset size and CNN performance. Tian et al. (2019) reported that performance improved with the number of fruit training images, reaching convergence around 3000 images. A similar analysis was performed by Koirala et al. (2019b) and Bargoti and Underwood (2017a), who reached convergence at around 400 training images and 500 000 annotated instances, while Wang et al. (2022) showed that 2500 annotated objects were sufficient for single-class fruit training.

Lu and Young (2020) reviewed publicly available datasets that could be of interest for training future fruit detection CNNs. Table 3.2 provides details of the ten datasets included in Lu and Young (2020) and seven additional datasets for fruit detection, classification and segmentation.

Table 3.2. Publicly available datasets for fruit detection. Data can be accessed by clicking on the corresponding title (highlighted in blue).

Title	Year	Image type	Images (instances) *	Image size **	Annotation type	Crops	References
ACFR-orchard fruit dataset	2016	RGB	3704	308×202/ 500×500	Bounding boxes	Almonds, apples, mangoes	(Bargoti and Underwood, 2017a)
DeepFruits	2016	RGB	586	Different sizes	Bounding boxes	7 different fruits	(Sa et al., 2016)
MangoNet semantic dataset	2018	RGB	49	4000×3000	Segmentation masks	Mangoes	(Kestur et al., 2019)
Date fruit dataset	2019	RGB Videos	8079 15	Different sizes	Length, weight, maturity	Dates	(Altaheri et al., 2019)
Embrapa WGISD	2019	RGB	300 (4432)	2048×1365/ 5184×3456	Instance segmentation	Grapes	(Santos et al., 2020)
ISARLab counting dataset	2019	RGB	1560	300×300/ 606×403	Fruit number per image	Almonds, olives, apples	(Bellocchio et al., 2019)
Kfuji-RGB-DS dataset	2019	RGB+Depth+NIR	967 (12839)	548×373	Bounding boxes	Apples	(Gené-Mola et al., 2019b)
MangoYOLO data set	2019	RGB	1730	612×512	Bounding boxes	Mangoes	(Koirala et al., 2019b)
MinneApple	2019	RGB	1000 (41000)	1280×720	Instance segmentation	Apples	(Häni et al., 2020b)
WSU apple dataset	2019	RGB	2298	Different sizes	Bounding boxes	Apples	(Bhusal et al., 2019)
Apple detect dataset	2020	RGB	5969	1024×1024	Apple centre point	Apples	(Biffi et al., 2021)
FruitsGB: Top Indian fruits with quality	2020	RGB	12000	256×256	Quality label	6 different fruits	(Meshram et al., 2020)
Fuji-SfM dataset	2020	RGB Point cloud	288 (1749) 1 (1455)	1024×1024 10.5 Mpts	Segmentation masks 3D bounding boxes	Apples	(Gené-Mola et al., 2020e)
LFuji-air dataset	2020	Point cloud	88 (1444)	235 kpts	3D bounding boxes	Apples	(Gené-Mola et al., 2020b)
Scifresh-apple-RGB-images	2020	RGB	800	1920×1080	Bounding boxes	Apples	(Gao et al., 2020)
Mango fruit on tree image collection	2021	RGB	250	4752×3168	Fruit number per image	Mangoes	(Walsh et al., 2021)
PFuji-Size dataset	2021	Point cloud	4 (615)	9.1 Mpts	3D instance segmentation + fruit centre location + diameters	Apples	(Gené-Mola et al., 2021b)

In the case of point cloud based datasets: *number of point clouds provided in the dataset, ** number of points per point cloud (average).

When a CNN model trained with a given dataset do not generalize well with new data, semi-automatic labelling is an option for better annotation efficiency. Semi-automatic labelling consists of automatically detecting fruits in new images with a pre-trained network (or an unsupervised method) and generating the ground truth by manually correcting the detections (dos Santos Ferreira et al., 2019). Another option is to use weakly supervised methods: Bellocchio et al. (2019, 2020) proposed a deep learning approach that only required a simple image binary labelling; Biffi et al. (2021) proposed a deep learning approach based on an adaptive training sample selection (ATSS) method that only requires annotation of the centre point of the objects; while Bhattarai and Karkee (2022) proposed a regression network (CountNet) which only requires the ground truth of the number of fruits per image.

When the number of empirical data is limited, different strategies have been applied to increase the capability of the network to generalize. Data augmentation techniques use annotated images to create new images by means of image transformations such as image flipping, rotation, and colour perturbations. This is a common practice employed in fruit detection works (Koirala et al., 2019a). Another option is to use synthetic data. Bresilla et al. (2019) generated synthetic images with random elliptic dark-green shapes (leaves) and light-green and light-red circles (fruit). More recently, the introduction of cycle generative adversarial networks has shown an improvement of the realism of synthetic images (Zhang et al., 2021a), increasing the performance of trained networks by more than 8 % (Barth et al., 2020).

3.3.6 Fruit tracking and counting

A common source of error when estimating fruit load is the double counting of fruit. The easiest method to prevent this is to acquire data along the orchard without overlap between consecutive frames (Bargoti and Underwood, 2017a; Apolo-Apolo et al., 2020a). However, since the ratio of visible to occluded fruit is not always constant, the use of multi-view approaches is sometimes required to increase fruit detectability (Hemming et al., 2014). Hence, to prevent double counting, fruit need to be tracked during scanning.

Two different strategies have been applied to track fruit across consecutive frames: video multi-object tracking (MOT) and the use of 3D data to locate the position of detections in the 3D space (Fig. 3.6). So far, the method most commonly used for video fruit tracking has been the Kalman filter (Anderson et al., 2021a; Itakura et al., 2021; Liu et al., 2018; Wang et al., 2019). Alternatively, Das et al. (2015) used optical flow, while Stein et al. (2016) used epipolar geometry by projecting the epipolar lines of the detected fruits centre onto consecutive frames. Recently, deep learning has demonstrated a good performance for solving MOT tasks (Dendorfer et al., 2020), being DeepSORT (Wojke et al., 2017) the deep learning-based tracking method most used for fruit counting in videos (Osman et al., 2021; Parico and Ahamed, 2021; Villacrés et al., 2023). A different approach was implemented in Roy and Isler (2016), who used calibrated cameras to register images through affine transformation. Similarly, Apolo-Apolo et al. (2020b) and Chen et al. (2019), applied affine transformations to build an orthomosaic of the entire orchard and subsequently detected fruits. Another method involves utilizing image stitching, as demonstrated by Zhang et al. (2022b), who used a SIFT-based image matching technique to form unique panoramic image of the captured fruit trees.

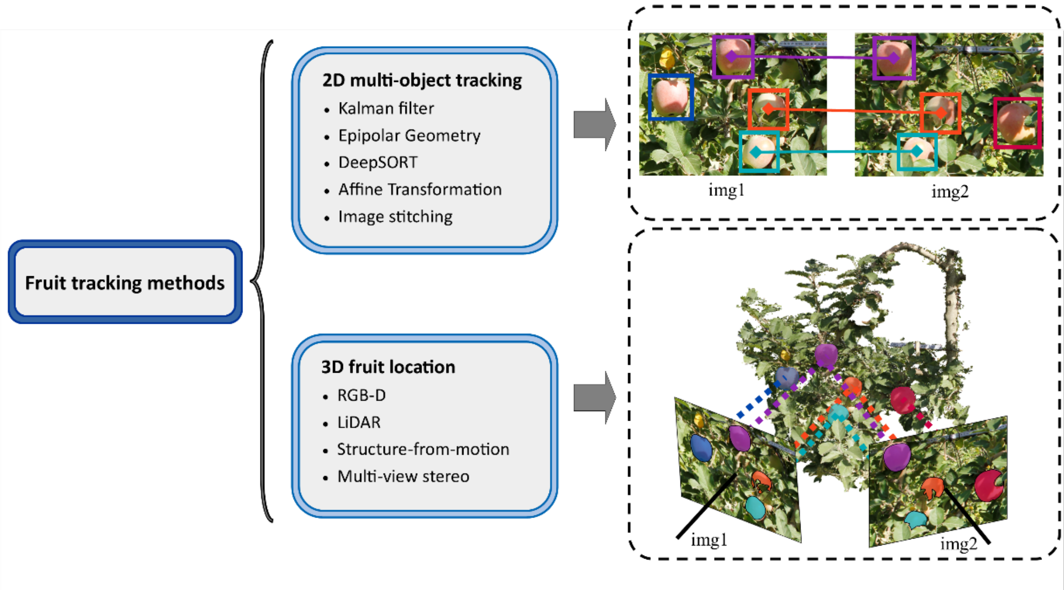


Fig. 3.6. Summary of methods used to prevent fruit double counting. Multi-object tracking (top) and 3D projection (bottom) procedures.

An alternative approach to reduce double-counting issues is the detection of fruit in the 3D space by means of RGB-D cameras, LiDAR sensors, or structure-from-motion (SfM). Wang et al. (2013) used a stereovision system synchronized with two global navigation satellite system (GNSS) receivers in order to transform apple locations into the global coordinate system. Then, fruit detected in consecutive frames closer than 0.16 m were automatically merged. Other works proposed the use of SfM to merge fruit detected from different camera positions (Gené-Mola et al., 2020d; Häni et al., 2020a; Liu et al., 2018, 2019; Santos et al., 2020). Taking advantage of the 3D data generated with SfM photogrammetry, fruits are previously detected in images and subsequently projected onto the 3D space for pair-wise association (Table 3.3).

Table 3.3. A comparative table of results reported in different fruit counting works. Results are reported in terms of R^2.

Tracking method	Sensors	Fruit detection method	Backbone	Crops	R^2 *	Reference
Images without overlap	RGB	MLP + CHT	N/A	Apples	0.83	(Bargoti and Underwood, 2017b)
Epipolar geometry	RGB + LiDAR	Faster-RCNN	VGG-16	Mangoes	0.90	(Stein et al., 2016)
Kalman	RGB	SSD	Mobilenet	Avocado, apples, lemons	0.77	(Vasconez et al., 2020)
	RGB	MangoYolo	Not specified	Mangoes	0.62 (DR)	(Wang et al., 2019)
DeepSORT	RGB	YOLOv4	CSPDarknet53	Pears	0.755 (MOTA)	(Parico and Ahamed, 2021)
Orthomosaic	RGB	Faster-RCNN	Resnet V2 Atrous	Apples	0.80	(Apolo-Apolo et al., 2020b)
	RGB	Faster-RCNN	Resnet-50	Strawberries	0.84 (DR)	(Chen et al., 2019)
3D projection	RGB	HSV thresholding	N/A	Apples	0.12 (ADRE)	(Wang et al., 2013)
	RGB	Faster-RCNN	Not specified	Mangoes	0.78	(Liu et al., 2019)
	RGB	Mask-RCNN	ResNet-101-FPN	Apples	0.80	(Gené-Mola et al., 2020d)

* Average detection rate error (ADRE), detection rate (DR) and MOT accuracy (MOTA) are provided when the R^2 coefficient is not available.

3.4 *Fruit size and maturity estimation*

3.4.1 Size estimation from 2D images

This group includes the set of works carried out by Stajnko et al., where apple fruit diameters were estimated throughout their growing season using RGB (Stajnko and Čmelik, 2005; Stajnko et al., 2009) and thermal images (Stajnko et al., 2004). A high coefficient of determination was obtained when comparing the estimated fruit diameter growing curves with the actual ones (R2 of 0.89 and 0.96 for RGB and thermal images, respectively). The tests with thermal cameras also showed that it is more difficult to detect the thermal gradient of the fruits inside the crown; this is because they heat up less than fruit located on the outside part. Likewise, Wang et al. (2020) used a spherical video camera for monitoring the apple growth from fruit thinning to their ripening. Estimates of the horizontal diameter of apples were made by applying ellipse and circular fitting methods and with the help of calibration balls. Ellipse fitting estimates yielded a mean average absolute error of 0.90 mm, much less than the 2.80 mm error obtained using a circular fitting. The size of citrus fruit was also estimated by Apolo-Apolo et al. (2020a), in this case using images taken from an unmanned aerial vehicle (UAV) and considering a wood ruler of known dimensions as the calibration object. Recently, Lu et al. (2022) proposed a near real-time apple fruit detection and sizing method from images taken by a low-cost smartphone in various growth stages. To estimate the fruit size, a red artificial apple was placed as a reference in the middle of the target area during the data collection stage. Estimated fruit sizes achieved R2 values of 0.68 and 0.66 in fruit height and fruit width, respectively.

Another alternative is based on knowing the distance to the camera of each of the fruit that appear in the image. In their pioneering work, Regunathan and Lee (2005) combined colour images with distance information obtained with ultrasound sensors and, using trigonometry, estimated the dimensions of citrus fruit. In this line, Wang et al. (2017) used the images and depth data provided by an RGB-D camera to estimate the size of mango fruits in trees by applying the thin lens theory. Root-mean-square errors (RMSE) of 4.9 and 4.3 mm were obtained in the fruit length and width estimates, respectively. Gongal et al. (2018) estimated apple sizes from distances provided by a time-of-flight (ToF) camera using a regression model that converts pixels (digital camera) to millimetres. The mean absolute percentage error (MAPE) was 15.2 %, lower than the 30.9 % obtained when the size was derived from the point clouds provided by a ToF camera. Another approach to estimate the fruit size from two images taken at different positions was presented by Rakun et al. (2019). This procedure uses image registration and similar triangles, known the distance between the two camera positions. Average diameter errors of 7 and 8 mm were obtained for peach and apple fruits, respectively.

3.4.2 Size estimation from 3D point clouds

As mentioned, fruit size estimation based on 2D images require the use of calibration targets or to merge the image data with ancillary distance information that adds complexity and computational costs to the processing. These limitations can be overcome using 3D sensing techniques (Rosell and Sanz, 2012; Gregorio and Llorens, 2021), such as LiDAR, structured-light, binocular stereo vision, multi-view stereo (MVS) or RGB-D cameras, among others, which allow the generation of three-dimensional reconstructions of the fruits.

Regarding LiDAR-based techniques, Méndez et al. (2019) used a 3D laser scanner with RGB data and applied the k-means algorithm to estimate the number and size of oranges. The computed diameters did not show significant differences in relation to those measured manually ($p=0.35$). As the authors point out, this is a time-consuming method, but given its high accuracy it can serve as a reference for other faster and more economical methods. For their part, Tsoulias et al. (2020) used a mobile terrestrial LiDAR scanner to monitor apples at different growth stages. Fruit diameter was estimated from each point cluster identified as apple and the resulting R2 with RMSE was 0.46 with 10.8 % and 0.67 with 7.7 %, 42 DAFB and at harvest, respectively.

Structured-light principle was used by Rist et al. (2019), who tested a hand-held high-resolution scanner for 3D phenotyping of grape bunches under field conditions. These are high cost, high precision devices with acquisition speeds of about 1 million points/s. The authors achieved R2 values of 0.70 to 0.91 in the prediction of several phenotypic traits (number and diameter of grapes; bunch width, length and volume). The RMSE values were 13.51 and 19.24 mm for bunch width and length, respectively, and 28.09 mL for the volume.

Binocular stereo vision was applied in harvesting robots under outdoor conditions by Luo et al. (2016). The authors proposed a method to detect the cut-off point of the peduncle and estimate the volume of grape bunches. As a result of their work, they obtained errors of less than 17 mm and 19 mm in bunch height and diameter, respectively. Herrero-Huerta et al. (2015) applied MVS for vineyard phenotyping and determined the grape bunch volume using an automatic method that fit a convex hull to the point cloud and a semi-automatic method that generated a CAD model. In both methods, similar coefficients of determination were obtained (0.76 and 0.77) when comparing the estimates with the actual bunch volumes. MVS was also applied in vineyards by Rose et al. (2016), who determined berry diameter by fitting spheres to point clouds.

Estimates were highly accurate with differences of about 2 mm with respect to manual measurements. Recent studies (Gené-Mola et al., 2021a; Grilli et al., 2021) have applied MVS and SfM to carry out in-field diameter estimation of apple fruit. Gené-Mola et al. (2021a) compared the performance of four different size estimation methods under several fruit visibility/occlusion levels (Fig. 3.7). The least squares method was concluded to be the most efficient in terms of computational cost, while the MAE ranged from 4.5 to 7.8 mm depending on the visibility. These errors were lower than those obtained with the largest segment method and similar to those obtained with the M-estimator sample and consensus (MSAC) method and template matching. For their part, Grilli et al. (2021) developed a procedure for on-tree automatic apple fruit counting and sizing using videos acquired with a smartphone. Apple size estimation was performed by fitting spheres (RANSAC method) on the point cloud.

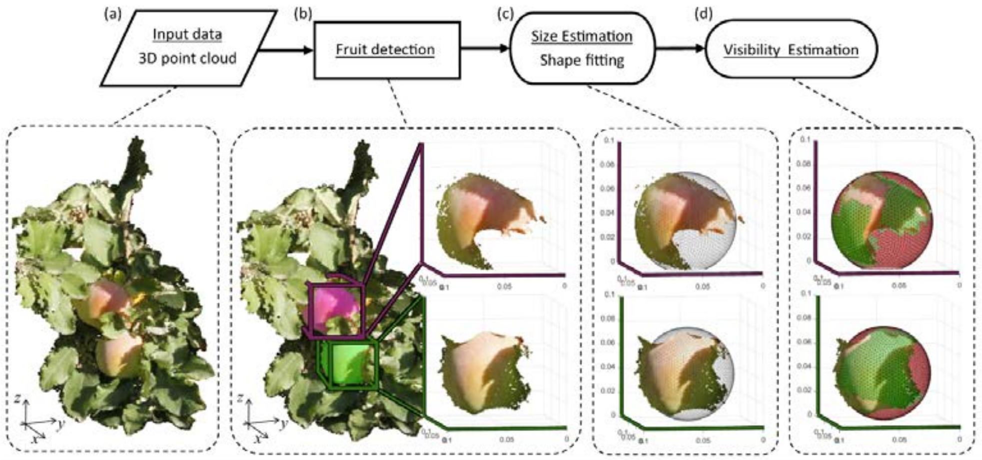


Fig. 3.7. Pipeline proposed by Gené-Mola et al. (2021a) to obtain in-field diameter estimation of apple fruit.

RGB-D sensors have been applied in vineyard yield estimation by Hacking et al. (2019). In their study, RGB-D measurements were used to create one mesh per grape bunch and determine its volume and mass. Also in vineyards, Kurtser et al. (2020b) used point clouds generated by an RGB-D camera to detect the grape clusters. These authors proposed three methods based on fitting geometric shapes to estimate the grape cluster size, obtaining the best results using percentile bounding boxes. Also using RGB-D sensors, Yu et al. (2022b) performed 3D sphere fitting to estimate the position and size of pomegranate fruits. Estimates of the fruit radius presented an RMSE of 2.35 mm and R2 of 0.826 when compared to the actual radius, while the position error was less than 5 mm.

As seen in this section, fruit size estimations have been carried out in a limited number of studies, many of them focusing on a few species. It is difficult to compare the performance of the different techniques due to the diversity of metrics used for their evaluation (Table 3.4). It is therefore advisable that future works include, at least, the mean absolute error (MAE) and the coefficient of determination (R2) when comparing estimated and actual size values.

3.4.3 Advancing fruit maturity estimation

In addition to the fruit size, knowing their maturity is essential for proper crop load management as well as for subsequent postharvest processes. Although automatic methods for fruit maturity estimation are less developed than sizing methods, some pioneering works have been carried out. Since many fruit species exhibit specific change of shape during fruit development, the fruit maturity can be estimated by means of the shape of the singularized, segmented fruit data. In apple, the shape of fruit was modelled by means of statistical approach (Danckaers et al., 2017) or Fourier signature (Rogge et al., 2015; Tapia Zapata et al., 2022). Such approaches provide the next step of extracting information describing the maturity of fruit. In mango, the change around the shoulder of the fruit in dicates maturity, which was analysed by means of RGB imaging (Sahu and Potdar, 2017).

Beside the shape of fruit also the pigment content and distribution provide information on the fruit maturity. The pigment contents have been addressed by means of colour analysis and spectral-optical data with enhanced resolution providing information on the reflectance intensity altered by absorption of pigments at their specific wavebands (Merzlyak et al., 2003; Walsh et al., 2020). Measurements were carried out in contact to the fruit to avoid stray light or with passive RGB sensors being not reliable in varying lighting conditions. However, the intensity measured by means of RGB-D and LiDAR sensors was employed previously to analyse the pigments of whole canopies employing LiDAR sensors emitting at 532 nm (green) or 660 nm (red), the latter to measure the leaf chlorophyll content (Eitel et al., 2010). Accordingly, 3D fruit segmentation and chlorophyll analysis were recently shown on apples in the orchard (Tsoulias et al., 2023) and banana fruit in postharvest (Saha and Zude-Sasse, 2022). Employing the return signal strength intensity of LiDAR sensor requests the radiometric calibration referencing the lowest and highest measurable intensity as well as curvature correction (Saha and Zude-Sasse, 2022). Classification of different measuring dates during fruit development were shown for apple as well as banana fruit, providing an interesting alternative to multispectral 2D readings.

Table 3.4. Sensing techniques and methods for in-field fruit size estimation reporting coefficient of determination (R^2), absolute error (AE), mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), mean average percentage error (MAPE).

Techniques	Size estimation method	Fruits	Size parameters	Performance	References
Binocular stereo vision	Calibration object. Regression model to predict pixel sizes.	Grape	Bunch diameter/ height	AE < 18.6 mm / 16.2 mm	(Luo et al., 2016)
High resolution 3D scanner	Sphere fitting.	Grape	Bunch length / width / volume	R^2 = 0.70 / 0.71 /0.91	(Rist et al., 2019)
LiDAR-based sensor	Sphere fitting.	Apple	Fruit diameter	R^2=0.38 - 0.95 RMSE: 4.1 % – 15.8 % MAE= 3.5 mm – 12.4 mm MBE=-10.7 mm – 7.5 mm.	(Tsoulias et al., 2020)
MVS	Pixel conversion. Image registration.	Apple Peach	Fruit diameter Fruit diameter	MAE=8 mm MAE=7 mm [1]	(Rakun et al., 2019)
	Convex hull. CAD generation.	Grapes	Bunch volume	R^2 = 0.77 / 0.76 (convex hull / CAD model)	(Herrero-Huerta et al., 2015)
MVS, SfM	Sphere fitting.	Apple	Fruit diameter	R^2 = 0.91 RMSE = 5.1 mm MAE=3.7 mm MBE = -1.9 mm MAPE=5.9 %	(Gené-Mola et al., 2021a)
RGB + ultrasonic sensor	Distances with ultrasonic sensors. Pixel conversion.	Apple	Fruit diameter	RMSE=0.4 cm.	(Regunathan and Lee, 2005)
RGB camera	Calibration object. Pixel conversion.	Avocado Mandarin Navel orange Apple Mango Mango	Fruit diameter Fruit diameter Fruit diameter Fruit diameter Fruit length Fruit width	RMSE=3.4 mm RMSE=3.8 mm RMSE=2.4 mm RMSE=2.0 mm RMSE=5.3 mm / 5.5 mm (controlled/ambient light) RMSE=3.7 mm / 4.6 mm. (controlled/ambient light)	(Wang et al., 2018b)
	Calibration object. Pixel conversion.	Apple	Fruit diameter growing curve	R^2=0.96	(Stajnko et al., 2009)
	Pixel conversion.	Apple	Fruit diameter	r = 0.55 – 0.80 (harvest stage)	(Stajnko and Čmelik, 2005)
	Calibration object. Pixel conversion. Ellipse/ circle fitting.	Apple	Fruit diameter	MAE=0.90 mm / 2.80 mm (ellipse/circle fitting)	(Wang et al., 2020)
	Calibration object. Pixel conversion.	Grapes	Berry diameter	R^2= 0.88	(Roscher et al., 2014)
RGB camera + ToF	Calibration object. Regression model to predict pixel sizes.	Apple	Fruit diameter	MAPE: 15.2 % (RGB) / 30.9 % (ToF)	(Gongal et al., 2018)
RGB-D camera	Bounding box/ ellipsoid/ cylinder fitting.	Grapes	Bunch length / width	MAE=~2.9 cm /~3.6 cm.	(Kurtser et al., 2020b)
	Pixel conversion units.	Mango	Fruit length / width	RMSE=4.9 mm / 4.3 mm.	(Wang et al., 2017)
	Sphere fitting.	Pomegranate	Fruit radius	RMSE=2.35 mm, R^2 = 0.826	(Yu et al., 2022b)
Thermal cameras	Pixel conversion units.	Apple	Fruit diameter growing curve	R^2 = 0.89	(Stajnko et al., 2004)

3.5 Discussion and future trends

3.5.1 The importance of data acquisition

Sensors are the first stage of detection/counting and size estimation of fruit and, thus, are critical for the performance of the entire process. Limitations of up-to-date available sensors are transferred to the obtained measurements which feed up the subsequent applied algorithms, thus limiting their effectiveness. Changes in environmental lighting affects RGB and RGB-D cameras' performance (Gené-Mola et al., 2020c; Fu et al., 2020). In addition, structured light sensors usually fail to characterize the contours, what is especially problematic in small objects, such as fruit. Also, in contours of objects, LiDAR's mixed pixels phenomenon (Sanz-Cortiella et al., 2011) leads to distorted points clouds and filtering is often required.

Some more advanced and affordable new sensors, which are expected to achieve great advances in this field, are being already tested for fruit detection and sizing. Thus, multi-beam as well as solid state LiDAR sensors are a great step forward. There are also LiDAR + RGB systems that allow obtaining coloured point clouds, although the correct colour assignment needs further improvement, especially in the outlines of small objects. The possibility of using two LiDAR systems at different wavelengths to simultaneously determine the normalized difference vegetation index (NDVI) of fruit in addition to the fruit number and size has recently been demonstrated (Tsoulias et al., 2023). Some companies have developed systems that merge different sensing principles (sensor fusion) with AI and post processing algorithms in the same product (Zheng et al., 2021). Also, the use of smartphones' embedded sensors (GNSS, RGB cameras, LiDAR …) allows much more compact systems with post processing capabilities in the same hardware.

Some ideas for next steps towards fruit's detecting and sizing systems can be outlined, such as the combination of multiple sensors, with the same or different sensing principles (sensor fusion). In addition, the application of stereo vision, SfM and MVS principles to thermal cameras can also be assessed, in order to obtain point clouds in the thermal range of the electromagnetic spectrum. The same can be applied to multispectral (MS) cameras, which are commercially available at affordable prices: systems similar to those based on RGB cameras but with MS cameras can be developed, allowing to obtain 2D images of fruits/trees/crops including IR and, from these, obtain 2D images of vegetative indices (VI), such as NDVI or the normalized difference red edge index (NDRE), among others. Likewise, progress can be made in obtaining 3D point clouds in the IR and in these VI. In the case of MS cameras with multiple optical objectives (one for each spectral band), it is necessary to address the effect on the measurements of not having a single common optical objective.

Apart from the sensors themselves, it is also necessary to delve into the measurement system as a whole: vehicle, supports and optimal location of the sensors (Xie et al., 2022). The GNSS receiver system associated with the sensors is also important, since the more precise it is, the better, because it affects the accuracy of the measurements, especially with regard to the location of the fruit, which must allow mapping the size of fruit both in the tree and in the plot.

Another aspect is the optimization of the resolution of the images and point clouds obtained by the sensors, so that they do not compromise the processing speed and allow progress towards real-time detection. Progresses must also be made in the implementation of systems that are increasingly plug and play, to facilitate their effective implementation in the sector, without the need to be an ICT expert. Finally, more studies are needed to know how external variables - apart from lighting conditions - such as temperature, dust, fog, vibrations etc., influence sensors' performance.

3.5.2 Fruit counting

Fruit growers still often use manual fruit counts on trees sampled within the plot to estimate orchard fruit yield. The fruit grower's experience helps to make this task more efficient, but since fruit counting is manual, it is always very laborious and expensive. AI application is expected to become the new paradigm in providing fruit growers with fast and reliable fruit counting methods for yield estimation. However, different strategies can be proposed, some of which still require new advances to be applied in a practical way. As demonstrated in this review, getting to automate fruit counting is not a simple matter, when occlusions, varying background, changing lighting exposure, unstructured canopies, and variable crop-load level are some of the challenges to face (Bhattarai and Karkee, 2022). If this were not enough, different steps must be addressed in the overall fruit counting process adding even more computational complexity. To give an example of the difficulty involved, it is known that only object-level annotation takes a substantial amount of manual annotation hours to create large labelled datasets (Pawara et al., 2020). The regression-based fruit counting approach has also been raised in some research (Bhattarai and Karkee, 2022; Pawara et al., 2020). In contrast to detection-based, annotation of only the total number of fruits at image level is used to train a neural network for counting. Thus, there is no need for explicit individual detection and localization resulting in a computationally simpler process. Another particularly interesting strategy is the one mentioned in Hobbs et al. (2021), where a deep learning-based density estimation approach is applied to count the number of flowering pineapple plants. By combining the latest advances in remote sensing and computer vision, counting is then affordable in orchards with high planting density. Indirect yield prediction, more than fruit count, has also been implemented for years by developing models that relate yield to features from environment (meteorological information) and/or features from canopy or tree physiology (management mode, plant growth state) (He et al., 2022).

The automatic fruit detection with computer vision algorithms is a key task for fruit counting systems. From 2016, the introduction of deep learning stablished object detection convolutional neural networks as the standard method to detect fruit in images, achieving F1-scores higher than 90 %, similar to the human eye. The tendency of the used CNN architectures is being the following. First (from 2016 to 2018, approximately), importance was given to improve accuracy. In this period, architectures such as Faster-RCNN demonstrated to be more accurate than the previous methods. Later (from 2018 to 2021, approximately), efforts were focused on improving efficiency and speed (Table 3.1). During this period, one-stage networks such as YOLO and its variance became the most popular, demonstrating real-time processing speeds and similar accuracy than the previous architectures. The current trend is to develop lightweight CNN to be implemented for edge computing purposes processed in embedded computers. It is expected that this will facilitate the deployment of commercial and affordable fruit counting devices (Zhang et al., 2021b).

Future research in fruit detection is expected to introduce emerging machine learning methods such as vision transformers (Carion et al., 2020), which are promising deep learning architectures based on attention mechanisms which could be more efficient and accurate than the popular CNNs. On the other hand, it is also expected an advance on point cloud-based object detection algorithms, which so far, have shown a lower performance compared with 2D algorithms based on CNNs. In this regard, further research should be done in order to test 3D machine learning methods that have not been applied for fruit detection such as the use of graph neural networks (Zhou et al., 2020) or PointPillars (Lang et al., 2019).

Based on the revised literature, authors consider the detection problem a quite mature problem at the level of computer vision. The handcrafted methods have been largely superseded by those based on deep learning. The lack of generalization in the detection of fruits together with variable conditions of the acquisition process (lighting changes, noise, background colours, etc.) are the main factors that directly affect handcrafted methods. Nevertheless, there are still environments (e.g. high-contrast fruit compared to the background) where methods based on handcrafted features could continue to be advantageous given their low computational cost. In order to advance to the development and deployment of commercial devices, future works should apply and evaluate the methods for in-field counting. Having high detection rates in the images does not ensure a high performance of these systems for yield estimation and mapping, since there are other factors that affect the systems performance such as the structure of trees, the amount of fruit occlusions, the use of multi-view methods, the strategies to prevent fruit double counting (such fruit tracking), etc. During the literature review we found that there are many works that evaluate the detection performance in images, but few works comparing the number of fruits detected in the images with respect to the actual number of fruits on trees or orchards. In addition, very few fruit counting and mapping methods were evaluated at different growth stages and different scanning conditions to ensure that systems generalize well at different environments. Thus, further efforts should be done to confront the challenges not related with the detection and evaluate the generalization of the models on larger datasets including different orchards scanned at different conditions.

Finally, authors involved in future research should also consider making publicly available the codes and the datasets with detailed explanations about how to implement and use them. This will facilitate that the scientific community advances efficiently and collaboratively. In other research fields such as Computer Science it is a common practice to make codes and data available. This, for example, explains the rapid advances in deep learning during the last decade. However, in the field of fruit detection and counting there is still a reluctance for the open science, which makes difficult to reproduce the methods and makes it difficult collaborative and additive research. Counting fruits is a task in which AI has allowed great advances. But, thinking of applying sensors and processes punctually within the plots, the combined use of AI together with efficient sampling cannot be ruled out, this being a still pending issue.

3.5.3 Fruit sizing and characterization

The fruit sizing task has not received as much attention as fruit detection, but several advances have been achieved during the last decade. Most of the revised works measure the fruit size in pixels in images and then apply a conversion from pixels to millimeters. Methods based on 2D images require the usage of calibration targets placed at the same distance to the cameras than fruit, which limits the efficiency of the data acquisition process. However, more advanced methods are based on 3D data, which can directly measure the fruit size in millimeters, or on RGB-D data, which allow the conversion from pixels to millimeters by applying the pinhole camera model.

So far, the dimensions of fruit at advanced ripening stages have been estimated, but there is also an interest in earlier maturity stages. For instance, the measurement of apple fruitlets is of interest for precisely adjusting the dose when applying chemical thinning. Thus, while accurate fruit sizing results have been reported in the revised literature, further research should be done for measuring young fruit to take appropriate actions in crop load management.

One of the major challenges when measuring fruit size with sensors is the presence of fruit occlusions. Although a high percentage of fruits are partially occluded, some fruit sizing works found in the literature limit the evaluation of their methods on fully visible fruits. From the authors' opinion, to transfer the fruit sizing research methods needs to deal with occlusions. Consequently, future works should provide fruit sizing results at different degrees of occlusion. In addition, further research should be carried out to automatically estimate the percentage of visibility of detected fruits, which would allow to identify the most occluded detections and limit the measurement to the most visible fruits, which are likely to be better measured.

The authors consider that future works will involve the development of methods capable of real-time monitoring the fruit temperature (sunburn risk), estimating fruit maturity (section 3.4.3), early detecting its defects and diseases, etc. Ultimately, current size estimation methods should evolve towards fruit characterization methods to allow a more complete knowledge of the different variables that affect fruit growth.

3.5.4 Opportunities for research transfer

Considering in-field fruit size estimation (necessary to monitor growth and estimate fruit weight), new research and commercial opportunities are emerging with the priority of developing robust and low-cost systems, and also under the premise of having to process large amounts of data when applied to large farms with large number of trees. Currently, there are a few companies that detect and count blossom and fruits and, in some cases, estimate fruit dimensions to estimate yield (Anderson et al., 2021b). Fruit detection and sizing would unlock the possibility to estimate per tree crop load and adjust the number of fruits on a tree and quality bases. It is well-known that crop load influences fruit quality (Serra et al., 2016; Embree et al., 2007). Together with crop load, irrigations strategies but also fruit location in the canopy, mainly according to height, are also affecting quality parameters (Alcobendas et al., 2013). In-field fruit location and sizing systems allow fruits to be georeferenced on a per tree basis and also to register their position in the canopy. An early detection and sizing solution would allow the farmers to apply thinning strategies within the same season considering the number of fruitlets per tree or even per branch or according to height. Several measurements along the season would provide him/her feedback about fruit growth uniformity and expected fruit quality. Late measurements, right before harvesting, would provide feedback on the applied strategy and information for the next season. When use in a whole farm approach instead of in a sampling approach, those systems would allow farms to better plan their logistics (labor force hiring, distribution within the field, transport and storage capacity, etc.) and also accurately estimate their yield and benefit according to fruit size classifications. In addition, when fruit size distribution is obtained for a whole plot, real-time or even map-based selective harvesting strategies could also be applied after a cost-benefit analysis according to fruit size or even to fruit colour (when such information is also gathered).

Summarizing, having information on the number of fruit, their size and their location within the trees and throughout the plot on a continuous, non-discreet, bases, will allow farms apply fruit quality and or cost-benefit -oriented strategies and make more informed decisions in the framework of Precision

Agriculture or Precision Fructiculture resulting in enhanced fruit quality and reduced fruit size distribution.

3.6 Conclusions

From the analysis of previous research in fruit detection and sizing, it can be concluded that, although very significant advances have been achieved in the recent past (in particular since the development of deep learning algorithms), it remains as an open field of study, which is currently a focal point of great interest.

Fruit load management and yield estimation in fruit orchards is still usually done by manual/visual fruit counting and sizing. However, this is always a costly task in terms of time and labour. For this reason, automatic counting using fruit detection systems is becoming a feasible option for the fruit sector.

Actually, both leaf area and fruit size can be estimated with LiDAR, RGB and RGB-D sensors-based systems, enabling the tree-individual analysis of fruit bearing capacity. Such precise management avoids errors and, therefore, can contribute to more sustainable fruit production. In postharvest, the fruit size determines the fruit value in some crops such as sweet cherry. In other crops such as apples, the storability of fruit can be affected by fruit size.

Both active and passive electromagnetic (EM) radiation-based sensors are being used for detection and sizing of fruits, most of them in the visible, IR or thermal region of the EM spectrum. Hopeful future advances are expected from new emerging sensor, electronics and post processing systems as well as sensor fusion, which should lead to achieving this goal in a practical and affordable way in a few years. Optimizing measurements' size files, GNNS accuracy and systems' simplicity of use no doubt will help greatly to the adoption of the commercially products that will gradually appear in the coming years.

Hand in hand with these advances in sensors, a key point has also been the developments made in the applied algorithms, taking special relevance those based on artificial intelligence techniques and specifically deep learning based on convolutional neural networks, CNNs. Most fruit detection and sizing recent approaches use image classification, object detection or semantic and instance segmentation CNNs. However, near future advances are also linked to the availability of high quality and size datasets to train the algorithms that will be developed from now on.

3.7 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) and by the Spanish Ministry of Science and Innovation / AEI/10.13039/501100011033 / FEDER (grants RTI2018-094222-B-I00 [PAgFRUIT project] and PID2021-126648OB-I00 [PAgPROTECT project]). The Secretariat of Universities and Research of the Department of Business and Knowledge of the Generalitat de Catalunya and European Social Fund (ESF) are also thanked 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.

3.8 References

Alcobendas, R., Mirás-Avalos, J.M., Alarcón, J.J., Nicolás, E., 2013. Effects of irrigation and fruit position on size, colour, firmness and sugar contents of fruits in a mid-late maturing peach cultivar. Sci. Hortic. 164(17), 340-347. https://doi.org/10.1016/j.scienta.2013.09.048.

Altaheri, H., Alsulaiman, M., Muhammad, G., Amin, S.U., Bencherif, M., Mekhtiche, M., 2019. Date fruit dataset for intelligent harvesting. Data Br. 26, 104514. https://doi.org/10.1016/j.dib.2019.104514

Anderson, N.T., Walsh, K.B., Koirala, A., Wang, Z., Amaral, M.H., Dickinson, G.R., Sinha, P., Robson, A.J., 2021a. Estimation of fruit load in Australian mango orchards using machine vision. Agronomy 11(9), 1711. https://doi.org/10.3390/agronomy11091711

Anderson, N.T., Walsh, K.B., Wulfsohn, D., 2021b. Technologies for forecasting tree fruit load and harvest timing—From ground, sky and time. Agronomy 11 (7), 1409. https://doi.org/10.3390/agronomy11071409

Apolo-Apolo, O.E., Martínez-Guanter, J., Egea, G., Raja, P., Pérez-Ruiz, M., 2020a. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. Eur. J. Agron. 115, 126030. https://doi.org/10.1016/j.eja.2020.126030

Apolo-Apolo, O.E., Pérez-Ruiz, M., Martínez-Guanter, J., Valente, J., 2020b. A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a deep learning technique. Front. Plant Sci. 11, 1086. https://doi.org/10.3389/fpls.2020.01086

Arnó, J., 2008. Una Investigación sobre la Variabilidad Intraparcelaria en Viña y el Uso de Sensores Láser en Viticultura de Precisión (Spanish). Ph.D. Thesis, Universitat de Lleida, Catalonia, Spain. https://www.tesisenred.net/handle/10803/8150#page=1

Bargoti, S., Underwood, J., 2017a. Deep fruit detection in orchards, in: 2017 IEEE Int. Conf. Robot. Autom. (ICRA), pp. 3626–3633. https://doi.org/10.1109/ICRA.2017.7989417

Bargoti, S., Underwood, J.P., 2017b. Image segmentation for fruit detection and yield estimation in apple orchards. J. F. Robot. 34, 1039–1060. https://doi.org/10.1002/rob.21699

Barth, R., Hemming, J., Van Henten, E.J., 2020. Optimising realism of synthetic images using cycle generative adversarial networks for improved part segmentation. Comput. Electron. Agric. 173, 105378. https://doi.org/10.1016/j.compag.2020.105378

Bavay, C., Symoneaux, R., Maître, I., Kuznetsova, A., Bruun Brockhoff, P., Mehinagic, E., 2013. Importance of fruit variability in the assessment of apple quality by sensory evaluation. Postharvest Biol. Technol. 77, 67-74. https://doi.org/10.1016/j.postharvbio.2012.11.005.

Bechar, A., Vigneault, C., 2016. Agricultural robots for field operations: Concepts and components. Biosyst. Eng. 149, 94–111. https://doi.org/10.1016/j.biosystemseng.2016.06.014

Bellocchio, E., Ciarfuglia, T.A., Costante, G., Valigi, P., 2019. Weakly supervised fruit counting for yield estimation using spatial consistency. IEEE Robot. Autom. Lett. 4, 2348–2355. https://doi.org/10.1109/LRA.2019.2903260

Bellocchio, E., Costante, G., Cascianelli, S., Fravolini, M.L., Valigi, P., 2020. Combining domain adaptation and spatial consistency for unseen fruits counting: a quasi-unsupervised approach. IEEE Robot. Autom. Lett. 5, 1079–1086. https://doi.org/10.1109/LRA.2020.2966398

Bhattarai, U., Karkee, M., 2022. A weakly-supervised approach for flower/fruit counting in apple orchards. Comput. Ind. 138, 103635. https://doi.org/10.1016/j.compind.2022.103635

Bhusal, S., Karkee, M., Zhang, Q., 2019. Apple Dataset Benchmark from Orchard Environment in Modern Fruiting Wall [WWW Document]. Res. Exch. https://doi.org/10.7273/000001752

Biffi, L.J., Mitishita, E., Liesenberg, V., Santos, A.A. dos, Gonçalves, D.N., Estrabis, N.V., Silva, J. de A., Osco, L.P., Ramos, A.P.M., Centeno, J.A.S., Schimalski, M.B., Rufato, L., Neto, S.L.R., Marcato Junior, J., Gonçalves, W.N., 2021. ATSS deep learning-based approach to detect apple fruits. Remote Sens. 2021, 13(1), 54. https://doi.org/10.3390/rs13010054

Bonany, J., Buehler, A., Carbó, J., Codarin, S., Donati, F., Echeverría, G., Egger, S., Guerra, W., Hilaire, C., Höller, I., Iglesias, I., Jesionkowska, K., Konopacka, D., Kruczynska, D., Martinelli, A., Pitiot, C., Sansavini, S., Etehr, R., Schoorl, F., 2013. Consumer eating quality acceptance of new apple varieties in different European countries. Food Qual. Prefer. 30(2), 250-259. https://doi.org/10.1016/j.foodqual.2013.06.004.

Bresilla, K., Perulli, G.D., Boini, A., Morandi, B., Corelli Grappadelli, L., Manfrini, L., 2019. Single-shot convolution neural networks for real-time fruit detection within the tree. Front. Plant Sci. 10, 611. https://doi.org/10.3389/fpls.2019.00611

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. arXiv:2005.12872. http://arxiv.org/abs/2005.12872.

Casagrande, E., Génard, M., Lurol, S., Charles, F., Plénet, D., Lescourret, F., 2021. A process-based model of nectarine quality development during pre- and post-harvest. Postharvest Biol. Technol. 175, 111458. https://doi.org/10.1016/j.postharvbio.2020.111458

Chaivivatrakul, S., Dailey, M.N., 2014. Texture-based fruit detection. Precis. Agric. 15, 662–683. https://doi.org/10.1007/s11119-014-9361-x

Chen, J., Wu, J., Wang, Z., Qiang, H., Cai, G., Tan, C., Zhao, C., 2021. Detecting ripe fruits under natural occlusion and illumination conditions. Comput. Electron. Agric. 190, 106450. https://doi.org/10.1016/j.compag.2021.106450

Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting apples and oranges with deep learning: a data-driven approach. IEEE Robot. Autom. Lett. 2, 781–788. https://doi.org/10.1109/LRA.2017.2651944

Chen, Y., Lee, W.S., Gan, H., Peres, N., Fraisse, C., Zhang, Y., He, Y., 2019. Strawberry yield prediction based on a deep neural network using high-resolution aerial orthoimages. Remote Sens. 11(13), 1584. https://doi.org/10.3390/rs11131584

Cheng, H., Damerow, L., Sun, Y., Blanke, M., 2017. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. J. Imaging 3(1), 6. https://doi.org/10.3390/jimaging3010006

Chu, P., Li, Z., Lammers, K., Lu, R., Liu, X., 2021. Deep learning-based apple detection using a suppression mask R-CNN. Pattern Recognit. Lett. 147, 206–211. https://doi.org/10.1016/j.patrec.2021.04.022

Danckaers, F., Huysmans,T., Van Dael, M., Verboven, P., Nicolaï,B., Sijbers, J., 2017. Building 3D statistical shape models of horticultural products. Food Bioproc. Tech. 10, 2100-2112. https://doi.org/10.1007/s11947-017-1979-z

Das, J., Cross, G., Qu, C., Makineni, A., Tokekar, P., Mulgaonkar, Y., Kumar, V., 2015. Devices, systems, and methods for automated monitoring enabling precision agriculture, in: 2015 IEEE International Conference on Automation Science and Engineering (CASE), pp. 462-469. https://doi.org/10.1109/CoASE.2015.7294123

Delong, M.J., Prange, K.R., Harrison, A.P., Embree, G.C., Nichols, S.D., Wright, A.H., 2006 The influence of crop-load, delayed cooling and storage atmosphere on post-storage quality of 'Honeycrisp'™ apples. J. Hortic. Sci. Biotechnol. 81(3), 391-396, https://doi.org/10.1080/14620316.2006.11512078

Dendorfer, P., Ošep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., Leal-Taixé, L., 2020. MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking. Int. J. Comput. Vis. 129, 845-881. https://doi.org/10.1007/s11263-020-01393-0

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255. https://doi.org/10.1109/CVPRW.2009.5206848

dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T., 2019. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. Comput. Electron. Agric. 165, 104963. https://doi.org/10.1016/j.compag.2019.104963

Eitel, J.U.H., Vierling, L.A., Long, D.S., 2010. Simultaneous measurements of plant structure and chlorophyll content in broadleaf saplings with a terrestrial laser scanner. Remote Sens. Environ. 114(10), 2229–2237. https://doi.org/10.1016/j.rse.2010.04.025

Eizentals, P., Oka, K., 2016. 3D pose estimation of green pepper fruit for automated harvesting. Comput. Electron. Agric. 128, 127–140. https://doi.org/10.1016/j.compag.2016.08.024

Embree, C.G., Myra, M.T.D., Nichols, D.S., Wright, A.H., 2007. Effect of blossom density and crop load on growth, fruit quality, and return bloom in ‘Honeycrisp’ apple. HortScience 42(7), 1622-1625. https://doi.org/10.21273/HORTSCI.42.7.1622

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

FAO, 2017. The future of food and agriculture – Trends and challenges. Rome. http://www.fao.org/3/i6583e/i6583e.pdf (accessed 16 September 2022).

Farjon, G., Krikeb, O., Hillel, A.B., Alchanatis, V., 2020. Detection and counting of flowers on apple trees for better chemical thinning decisions. Precis. Agric. 21(3), 503–521. https://doi.org/10.1007/s11119-019-09679-1

Favre, L., Hunter, D.A., O’Donoghue, E.M., Erridge, Z.A., Napier, N.J., Somerfield, S.D., Hunt, M., McGhie, T.K., Cooney, J.M., Saei, A., Chen, R.K.Y., McKenzie, M.J., Brewster, D., Martin, H., Punter, M., Carr, B., Tattersall, A., Johnston, J.W., Gibon, Y., Heyes, J.A., Lill, R.E., Brummell, D.A., 2022. Integrated multi-omic analysis of fruit maturity identifies biomarkers with drastic abundance shifts spanning the harvest period in 'Royal Gala' apple. Postharvest Biol. Technol. 193, 112059. https://doi.org/10.1016/j.postharvbio.2022.112059

Fraser, D.G., Jordan, R.B., Künnemeyer, R., McGlone, V.A., 2003. Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. Postharvest Biol. Technol. 27(2), 185–196. https://doi.org/10.1016/S0925-5214(02)00058-3

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

Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y., Al-Mallahi, A., Zhang, Q., Li, R., Cui, Y., 2021. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. Precis. Agric. 22, 754-776. https://doi.org/10.1007/s11119-020-09754-y

Fu, L., Liu, Z., Majeed, Y., Cui, Y., 2018. Kiwifruit yield estimation using image processing by an Android mobile phone. IFAC-PapersOnLine 51(17), 185–190. https://doi.org/10.1016/j.ifacol.2018.08.137

Fu, L., Wu, F., Zou, X., Jiang, Y., Lin, J., Yang, Z., Duan, J., 2022. Fast detection of banana bunches and stalks in the natural environment based on deep learning. Comput. Electron. Agric. 194, 106800. https://doi.org/https://doi.org/10.1016/j.compag.2022.106800

Gai, R., Chen, N., Yuan, H., 2021. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. Neural Comput. Appl. https://doi.org/10.1007/s00521-021-06029-z

Gan, H., Lee, W.S., Alchanatis, V., Ehsani, R., Schueller, J.K., 2018. Immature green citrus fruit detection using color and thermal images. Comput. Electron. Agric. 152, 117–125. https://doi.org/10.1016/j.compag.2018.07.011

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M., Zhang, Q., 2020. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. Comput. Electron. Agric. 176, 105634. https://doi.org/10.1016/j.compag.2020.105634

Gené-Mola, J., Gregorio, E., Auat Cheein, F., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A., Rosell-Polo, J.R., 2020a. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. Comput. Electron. Agric. 168, 105121. https://doi.org/10.1016/j.compag.2019.105121

Gené-Mola, J., Gregorio, E., Auat Cheein, F., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A., Rosell-Polo, J.R., 2020b. LFuji-air dataset: Annotated 3D LiDAR point clouds of Fuji apple trees for fruit detection scanned under different forced air flow conditions. Data Br. 29, 105248. https://doi.org/10.1016/j.dib.2020.105248

Gené-Mola, J., Gregorio, E., Guevara, J., Auat, F., Sanz-Cortiella, R., Escolà, A., Llorens, J., Morros, J.R., Ruiz-Hidalgo, J., Vilaplana, V., Rosell-Polo, J.R., 2019a. Fruit detection in an apple orchard using a mobile terrestrial laser scanner. Biosyst. Eng. 187, 171–184. https://doi.org/10.1016/j.biosystemseng.2019.08.017

Gené-Mola, J., Llorens, J., Rosell-Polo, J. R., Gregorio, E., Arnó, J., Solanelles, F., Martínez-Casasnovas, J. A., Escolà, A., 2020c. Assessing the performance of RGB-D sensors for 3d fruit crop canopy characterization under different operating and lighting conditions. Sensors 20(24), 7072. https://doi.org/10.3390/s20247072

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J. R., Escolà, A., Gregorio, E., 2021a. 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., Sanz-Cortiella, R., Rosell-Polo, J.R., Escolà, A., Gregorio, E., 2021b. PFuji-Size dataset: A collection of images and photogrammetry-derived 3D point clouds with ground truth annotations for Fuji apple detection and size estimation in field conditions. Data Br. 39, 107629. https://doi.org/10.1016/j.dib.2021.107629

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J.R., Morros, J.-R.R., Ruiz-Hidalgo, J., Vilaplana, V., Gregorio, E., 2020d. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. Comput. Electron. Agric. 169, 105165. https://doi.org/https://doi.org/10.1016/j.compag.2019.105165

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J.R., Morros, J.R., Ruiz-Hidalgo, J., Vilaplana, V., Gregorio, E., 2020e. Fuji-SfM dataset: A collection of annotated images and point clouds for Fuji apple detection and location using structure-from-motion photogrammetry. Data Br. 30, 105591. https://doi.org/10.1016/j.dib.2020.105591

Gené-Mola, J., Vilaplana, V., Rosell-Polo, J.R., Morros, J.-R., Ruiz-Hidalgo, J., Gregorio, E., 2019b. KFuji RGB-DS database: Fuji apple multi-modal images for fruit detection with color, depth and range-corrected IR data. Data Br. 25, 104289. https://doi.org/10.1016/j.dib.2019.104289

Gené-Mola, J., Vilaplana, V., Rosell-Polo, J.R., Morros, J.-R.R., Ruiz-Hidalgo, J., Gregorio, E., 2019c. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. Comput. Electron. Agric. 162, 689–698. https://doi.org/10.1016/j.compag.2019.05.016

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., Lewis, K, 2015. Sensors and systems for fruit detection and localization: A review. Comput. Electron. Agric. 116, 8–19. https://doi.org/10.1016/j.compag.2015.05.021

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, Cham, pp. 59–92. https://doi.org/10.1007/978-3-030-78431-7_3

Grilli, E., Battisti, R., Remondino, F., 2021. An advanced photogrammetric solution to measure Apples. Remote Sens. 13(1), 3960. https://doi.org/10.3390/rs13193960

Grilli, E., Menna, F., Remondino, F., 2017. A review of point clouds segmentation and classification algorithms, in: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(2W3), pp. 339–344. https://doi.org/10.5194/isprs-archives-XLII-2-W3-339-2017

Habib, M.T., Majumder, A., Jakaria, A.Z.M., Akter, M., Uddin, M.S., Ahmed, F., 2020. Machine vision based papaya disease recognition. J. King Saud Univ. - Comput. Inf. 32(3), 300–309. https://doi.org/10.1016/j.jksuci.2018.06.006

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

Hampson, C., Sanford, K., Cline, J., 2002. Preferences of Canadian consumers for apple fruit size. Can. J. Plant Sci. 82(1), 165-167. https://doi.org/10.4141/P01-059

Häni, N., Roy, P., Isler, V., 2020a. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. J. F. Robot. 37, 263-282. https://doi.org/10.1002/rob.21902

Häni, N., Roy, P., Isler, V., 2020b. MinneApple: A benchmark dataset for apple detection and segmentation. IEEE Robot. Autom. Lett. 5(2), 852–858. https://doi.org/10.1109/LRA.2020.2965061

Harker, F.R., Gunson, F.A., Jaeger, S.R., 2003. The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. Postharvest Biol. Technol. 28(3), 333-347. https://doi.org/10.1016/S0925-5214(02)00215-6

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, in: 2017 IEEE Int. Conf. Comput. Vis (ICCV), pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322

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

He, L., Zhou, J., Zhang, Q., Karkee, M., 2015. Evaluation of multipass mechanical harvesting on ‘Skeena’ sweet cherries trained to Y-trellis. HortScience 50(8), 1178-1182. https://doi.org/10.21273/HORTSCI.50.8.1178

Heim, S., Zude-Sasse, M., 2014. Analyse der Fruchtwachstumsrate für die Bewertung einer selektiven Ernte bei Prunus avium 'Celeste' und Prunus domestica 'Tophit plus'. Annual conference DGG (Deutsche Gartenbauwissenschaftliche Gesellschaft), Dresden, Germany.

Hemming, J., Ruizendaal, J., Willem Hofstee, J., van Henten, E.J., 2014. Fruit detectability analysis for different camera positions in sweet-pepper. Sensors 14, 6032–6044. https://doi.org/10.3390/s140406032

Herrero-Huerta, M., González-Aguilera, D., Rodriguez-Gonzalvez, P., Hernández-López, D., 2015. Vineyard yield estimation by automatic 3D bunch modelling in field conditions. Comput. Electron. Agric. 110, 17–26. https://doi.org/10.1016/j.compag.2014.10.003

Hobbs, J., Prakash, P., Paull, R., Hovhannisyan, H., Markowicz, B., Rose, G., 2021. Large-scale counting and localization of pineapple inflorescence through deep density-estimation. Front. Plant Sci. 11, 599705. https://www.frontiersin.org/articles/10.3389/fpls.2020.599705/full

Itakura, K., Narita, Y., Noaki, S., Hosoi, F., 2021. Automatic pear and apple detection by videos using deep learning and a Kalman filter. OSA Contin. 4, 1688-1695. https://doi.org/10.1364/osac.424583

Iwanami, H., 2011. Breeding for fruit quality in apple. in: Jenks, M.A, Bebeli, P. (Eds.), Breeding for Fruit Quality, pp. 173-200. http:/doi.org/10.1002/9780470959350.ch8

Jarvinen, T.D., Choi, D., Heinemann, P., Schupp, J., Baugher, T.A., 2019. Tree trunk position estimation for accurate fruit counts in apple yield mapping, in: 2019 ASABE International Meeting, 1900918. https://doi.org/10.13031/aim.201900918

Jia, W., Zhang, Z., Shao, W., Hou, S., Ji, Z., Liu, G., Yin, X., 2021. FoveaMask: A fast and accurate deep learning model for green fruit instance segmentation. Comput. Electron. Agric. 191, 106488. https://doi.org/10.1016/j.compag.2021.106488

Kang, H., Chen, C., 2019. Fruit detection and segmentation for apple harvesting using visual sensor in orchards. Sensors 19, 4599. https://doi.org/10.3390/s19204599

Kang, H., Chen, C., 2020. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. Comput. Electron. Agric. 171, 105302. https://doi.org/10.1016/j.compag.2020.105302

Kelman, E., Linker, R., 2014. Vision-based localisation of mature apples in tree images using convexity. Biosyst. Eng. 118(1), 174–185. https://doi.org/10.1016/j.biosystemseng.2013.11.007

Kestur, R., Meduri, A., Narasipura, O., 2019. MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. Eng. Appl. Artif. Intell. 77, 59–69. https://doi.org/10.1016/j.engappai.2018.09.011

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019a. Deep learning – method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric. 162, 219–234. https://doi.org/10.1016/j.compag.2019.04.017

Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C., 2019b. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO'. Precis. Agric. 20(6), 1107-1135. https://doi.org/10.1007/s11119-019-09642-0

Kurtser, P., Ringdahl, O., Rotstein, N., Andreasson, H., 2020a. PointNet and geometric reasoning for detection of grape vines from single frame RGB-D data in outdoor conditions, in: Proc. North. Light. Deep Learn. Work. 1, pp. 1-6.

Kurtser, P., Ringdahl, O., Rotstein, N., Berenstein, R., Edan, Y., 2020b. 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

Lang, A.H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O., 2019. Pointpillars: Fast encoders for object detection from point clouds. arXiv 1812.05784. https://arxiv.org/abs/1812.05784

LeCun, Y., Bengio, Y. Hinton, G., 2015. Deep learning. Nature 521, 436-444. https://doi.org/10.1038/nature14539

Lee, J., Mattheis, J.P., Rudell, D.R., 2013. Fruit size affects physiological attributes and storage disorders in cold-stored 'Royal Gala' apples. HortScience 48(12), 1518-1524. https://doi.org/10.21273/HORTSCI.48.12.1518

Li, H., Lee, W.S., Wang, K., 2016. Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images. Precis. Agric. 17(6), 678–697. https://doi.org/10.1007/s11119-016-9443-z

Lin, G., Tang, Y., Zou, X., Li, J., Xiong, J., 2019. In-field citrus detection and localisation based on RGB-D image analysis. Biosyst. Eng. 186, 34–44. https://doi.org/10.1016/j.biosystemseng.2019.06.019

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context, in: European Conference on Computer Vision, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, T.H., Ehsani, R., Toudeshki, A., Zou, X.J., Wang, H.J., 2019. Identifying immature and mature pomelo fruits in trees by elliptical model fitting in the Cr–Cb color space. Precis. Agric. 20(1), 138–156. https://doi.org/10.1007/s11119-018-9586-1

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. SSD: Single shot multibox detector, in: European Conference on Computer Vision, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Liu, X., Chen, S.W., Aditya, S., Sivakumar, N., Dcunha, S., Qu, C., Taylor, C.J., Das, J., Kumar, V., 2018. Robust fruit counting: Combining deep learning, tracking, and structure from motion, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1045–1052. https://doi.org/10.1109/IROS.2018.8594239

Liu, X., Chen, S.W., Liu, C., Shivakumar, S.S., Das, J., Taylor, C.J., Underwood, J., Kumar, V., 2019. Monocular camera based fruit counting and mapping with semantic data association. IEEE Robot. Autom. Lett. 4(3), 2296–2303. https://doi.org/10.1109/LRA.2019.2901987

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Lu, S., Chen, W., Zhang, X., Karkee, M., 2022. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. Comput. Electron. Agric. 193, 106696. https://doi.org/10.1016/j.compag.2022.106696

Lu, Y., Young, S., 2020. A survey of public datasets for computer vision tasks in precision agriculture. Comput. Electron. Agric. 178, 105760. https://doi.org/10.1016/j.compag.2020.105760

Luo, L., Tang, Y., Zou, X., Ye, M., Feng, W., Li, G., 2016. Vision-based extraction of spatial information in grape clusters for harvesting robots. Biosyst. Eng. 151, 90–104. https://doi.org/10.1016/j.biosystemseng.2016.08.026

Maldonado, W., Barbosa, J.C., 2016. Automatic green fruit counting in orange trees using digital images. Comput. Electron. Agric. 127, 572–581. https://doi.org/10.1016/j.compag.2016.07.023

Mekhalfi, M.L., Nicolò, C., Ianniello, I., Calamita, F., Goller, R., Barazzuol, M., Melgani, F., 2020. Vision system for automatic on-tree kiwifruit counting and yield estimation. Sensors 20(15), 4214. https://doi.org/10.3390/s20154214

Méndez, V., Pérez-Romero, A., Sola-Guirado, R., Miranda-Fuentes, A., Manzano-Agugliaro, F., Zapata-Sierra, A., Rodríguez-Lizana, A., 2019. In-field estimation of orange number and size by 3D laser scanning. Agronomy 9(12), 885. https://doi.org/10.3390/agronomy9120885

Merzlyak, M.N., Solovchenko, A.E., Gitelson, A.A., 2003. Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. Postharvest Biol. Technol. 27(2), 88–103. https://doi.org/10.1016/S0925-5214(02)00066-2

Meshram, V., Thanomliang, K., Ruangkan, S., Chumchu, P., Patil, K., 2020. FruitsGB: Top Indian fruits with quality [WWW Document]. IEEE DataPort. https://dx.doi.org/10.21227/gzkn-f379

Miller, W.M., Zude-Sasse, M., 2004. NIR-based sensing to measure soluble solids content of Florida citrus. Appl. Eng. Agric. 20(3), 321-327. https://doi.org/10.13031/2013.16052

Musacchi, S., Serra, S., 2018. Apple fruit quality: Overview on pre-harvest factors. Sci. Hortic. 234, 409–430. https://doi.org/10.1016/j.scienta.2017.12.057.

Nanni, L., Ghidoni, S., Brahnam, S., 2017. Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognit. 71, 158-172. https:// doi.org/10.1016/j.patcog.2017.05.025

Nguyen, T.T., Vandevoorde, K., Wouters, N., Kayacan, E., De Baerdemaeker, J.G., Saeys, W., 2016. Detection of red and bicoloured apples on tree with an RGB-D camera. Biosyst. Eng. 146, 33–44. https://doi.org/10.1016/j.biosystemseng.2016.01.007

Nicastro, R., Carillo, P., 2021. Food loss and waste prevention strategies from farm to fork. Sustainability, 13(10), 5443. https://doi.org/10.3390/su13105443

Nyarko, E. K., Vidović, I., Radočaj, K., Cupec, R., 2018. A nearest neighbor approach for fruit recognition in RGB-D images based on detection of convex surfaces. Expert Syst. Appl. 114, 454–466. https://doi.org/10.1016/j.eswa.2018.07.048

OECD, 2021. Apples, International Standards for Fruit and Vegetables. OECD Publishing, Paris. https://doi.org/10.1787/12ebba9f-en-fr.

Osman, Y., Dennis, R., Elgazzar, K., 2021. Yield estimation and visualization solution for precision agriculture. Sensors 21(19), 6657. https://doi.org/10.3390/s21196657

Parico, A.I.B., Ahamed, T., 2021. Real time pear fruit detection and counting using YOLOv4 models and Deep SORT. Sensors 21(14), 4803. https://doi.org/10.3390/s21144803

Paul, V., Pandey, R., 2014. Role of internal atmosphere on fruit ripening and storability - A review. J. Food Sci. Technol. 51, 1223-1250. https://doi.org/10.1007/s13197-011-0583-x

Pawara, P., Boshchenko, A., Schomaker, L.R.B., Wiering, M.A., 2020. Deep learning with data augmentation for fruit counting, in Rutkowski, L. et al. (Eds.), Artificial Intelligence and Soft Computing, ICAISC 2020, pp. 203-214. https://link.springer.com/chapter/10.1007/978-3-030-61401-0_20

Pedraza, I.L.A., Diaz, J.F.A., Pinto, R.M., Becker, M., Tronco, M.L., 2019. Sweet citrus fruit detection in thermal images using fuzzy image processing, in: IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI 2019), pp. 182–193. https://doi.org/10.1007/978-3-030-36211-9_15

Penzel, M., Lakso, A.N., Tsoulias, N., Zude-Sasse, M., 2020. Carbon consumption of developing fruit and the fruit bearing capacity of individual RoHo 3615 and Pinova apple trees. Int. Agrophys. 34(4), 409–423. https://doi.org/10.31545/intagr/127540

Qi, C.R., Liu, W., Wu, C., Su, H., Guibas, L.J., 2018. Frustum PointNets for 3D object detection from RGB-D Data, in: 31st IEEE Conference on Computer Vision and PatternRecognition (CVPR 2018), pp. 918-927. https://doi:10.1109/CVPR.2018.00102.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: Deep learning on point sets for 3D classification and segmentation, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 77-85. https://doi.org/10.1109/CVPR.2017.16

Qian, J., Xing, B., Wu, X., Chen, M., Wang, Y., 2018. A smartphone-based apple yield estimation application using imaging features and the ANN method in mature period. Sci. Agric. 75(4), 273–280. https://doi.org/10.1590/1678-992x-2016-0152

Qureshi, W.S., Payne, A., Walsh, K.B., Linker, R., Cohen, O., Dailey, M.N., 2017. Machine vision for counting fruit on mango tree canopies. Precis. Agric. 18(2), 224–244. https://doi.org/10.1007/s11119-016-9458-5

Rakun, J., Stajnko, D., Zazula, D., 2019. Plant size estimation based on the construction of high-density corresponding points using image registration. Comput. Electron. Agric. 157, 288–304. https://doi.org/10.1016/j.compag.2018.12.053

Redmon, J., Farhadi, A., 2018. YOLOv3: An incremental improvement. Tech Report, arXiv1804.02767. http://arxiv.org/abs/1804.02767

Regunathan, M., Lee, W.S., 2005. Citrus fruit identification and size determination using machine vision and ultrasonic sensors, in 2005 ASAE Annual International Meeting, 053017. https://doi.org/10.13031/2013.19821

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rist, F., Gabriel, D., Mack, J., Steinhage, V., Töpfer, R., Herzog, K., 2019. Combination of an automated 3D field phenotyping workflow and predictive modelling for high-throughput and non-invasive phenotyping of grape bunches. Remote Sens. 11(24), 2953. https://doi.org/10.3390/rs11242953

Robinson T.L., Lakso A.N., Greene D., 2017. Precision crop load management: The practical implementation of physiological models. Acta Hortic. 1177, 381-390. https://doi.org/10.17660/ActaHortic.2017.1177.55

Rogge, S., Defraeye, T., Herremans, E., Verboven, P., Nicolaï, B.M., 2015. A 3D contour based geometrical model generator for complex-shaped horticultural products. J. Food Eng. 157, 24-32. https://doi.org/10.1016/j.jfoodeng.2015.02.006.

Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R., Förstner, W., 2014. Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. Comput. Electron. Agric. 100, 148–158. https://doi.org/10.1016/j.compag.2013.11.008

Rose, J.C., Kicherer, A., Wieland, M., Klingbeil, L., Töpfer, R., Kuhlmann, H., 2016. Towards automated large-scale 3D phenotyping of vineyards under field conditions. Sensors 16(12), 2136. https://doi.org/10.3390/s16122136

Rosell, J.R., Sanz, R., 2012. A review of methods and applications of the geometric characterization of tree crops in agricultural activities. Comput. Electron. Agric. 81, 124–141. https://doi.org/10.1016/j.compag.2011.09.007

Roy, A.M., Bhaduri, J., 2022. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. Comput. Electron. Agric. 193, 106694. https://doi.org/10.1016/j.compag.2022.106694

Roy, P., Isler, V., 2016. Surveying apple orchards with a monocular vision system, in: IEEE International Conference on Automation Science and Engineering, pp. 916-921. https://doi.org/10.1109/COASE.2016.7743500

Roy, P., Kislay, A., Plonski, P.A., Luby, J., Isler, V., 2019. Vision-based preharvest yield mapping for apple orchards. Comput. Electron. Agric. 164, 104897. https://doi.org/10.1016/j.compag.2019.104897

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. DeepFruits: A fruit detection system using deep neural networks. Sensors 16(8), 1222. https://doi.org/10.3390/s16081222

Saha, K.K., Zude-Sasse, M., 2022. Estimation of chlorophyll content in banana during shelf life using LiDAR laser scanner. Postharvest Biol. Technol. 192, 112011. https://doi.org/10.1016/j.postharvbio.2022.112011

Sahu, D., Potdar, R.M., 2017. Defect identification and maturity detection of mango fruits using image analysis. Am. J. Artif. Intell. l(1), 5-14. https://doi.org/10.11648/j.ajai.20170101.12

Saitone, T. L., Sexton, R.J., 2017. Agri-food supply chain: Evolution and performance with conflicting consumer and societal demands. Eur. Rev. Agric. Econ. 44, 634–657. https://doi.org/10.1093/erae/jbx003.

Santos, T.T., de Souza, L.L., dos Santos, A.A., Avila, S., 2020. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. Comput. Electron. Agric. 170, 105247. https://doi.org/10.1016/j.compag.2020.105247

Sanz-Cortiella, R., Llorens-Calveras, J., Rosell-Polo, J.R., Gregorio-Lopez, E., Palacin-Roca, J., 2011. Characterisation of the LMS200 laser beam under the influence of blockage aurfaces. Influence on 3D acanning of tree orchards. Sensors 11, 2751–2772. https://doi.org/10.3390/s110302751

Serra, S., Leisso, R., Giordani, L., Kalcsits, L., Musacchi, S., 2016. Crop load influences fruit quality, nutritional balance, and return bloom in ‘Honeycrisp’ apple. HortScience 51(3), 236-244. https://doi.org/10.21273/HORTSCI.51.3.236

Shi, G., Ranjan, R., Khot, L.R., 2020. Robust image processing algorithm for computational resource limited smart apple sunburn sensing system. Inf. Process. Agric. 7(2), 212–222. https://doi.org/10.1016/j.inpa.2019.09.007

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556. http://arxiv.org/abs/1409.1556

Stajnko, D., Čmelik, Z., 2005. Modelling of apple fruit growth by application of image analysis. Agric. Conspec. Sci. 70(2), 59–64.

Stajnko, D., Lakota, M., Hočevar, M., 2004. Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging. Comput. Electron. Agric., 42(1), 31–42. https://doi.org/10.1016/S0168-1699(03)00086-3

Stajnko, D., Rakun, J., Blanke, M., 2009. Modelling apple fruit yield using image analysis for fruit colour, shape and texture. Eur. J. Hortic. Sci. 74(6), 260–267.

Stein, M., Bargoti, S., Underwood, J., 2016. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors 16, 1915. https://doi.org/10.3390/s16111915

Sun, C., Aernouts, B., Saeys, W., 2021. Effects of harvest time, fruit size and cultivar on the bulk optical properties of Satsuma mandarin. Postharvest Biol. Technol. 175, 111412. https://doi.org/10.1016/j.postharvbio.2020.111412

Sun, Q., Chai, X., Zeng, Z., Zhou, G., Sun, T., 2022. Noise-tolerant RGB-D feature fusion network for outdoor fruit detection. Comput. Electron. Agric. 198, 107034. https://doi.org/10.1016/j.compag.2022.107034

Sun, S., Jiang, M., He, D., Long, Y., Song, H., 2019. Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm. Biosyst. Eng. 187(22), 201–213. https://doi.org/10.1016/j.biosystemseng.2019.09.006

Syal, A., Garg, D., Sharma, S., 2014. Apple fruit detection and counting using computer vision techniques, in: 2014 IEEE International Conference on Computational Intelligence and Computing Research (IEEE ICCIC 2014), pp. 405–412. https://doi.org/10.1109/ICCIC.2014.7238364

Tao, Y., Zhou, J., 2017. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. Comput. Electron. Agric. 142, 388–396. https://doi.org/10.1016/j.compag.2017.09.019

Tapia Zapata, N., Saha, K.K., Tsoulias, N., Zude-Sasse, M., 2022. Fourier analysis of LiDAR scanned 3D point cloud data for surface reconstruction and fruit size estimation, in: 2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), pp. 197-202. https://doi.org/10.1109/MetroAgriFor55389.2022.9964881

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z., 2019. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agric. 157, 417–426. https://doi.org/10.1016/j.compag.2019.01.012

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

Tsoulias, N., Saha, K.K., Zude-Sasse, M., 2023. In-situ fruit analysis by means of LiDAR 3D point cloud of normalized difference vegetation index (NDVI). Comput. Electron. Agric. 205, 107611. https://doi.org/10.1016/j.compag.2022.107611

Tsoulias, N., Xanthopoulos, G., Fountas, S., Zude-Sasse, M., 2022. Effects of soil ECa and LiDAR-derived leaf area on yield and fruit quality in apple production. Biosyst. Eng. 223(B), 182-199. https://doi.org/10.1016/j.biosystemseng.2022.03.007

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H., Xue, Y., 2020. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. Precis. Agric. 21, 1072–1091. https://doi.org/10.1007/s11119-020-09709-3

UN, 2022. World population prospects 2022. Department of Economic and Social Affairs. Population Division. https://population.un.org/wpp/_(accessed 16 September 2022).

Underwood, J.P., Hung, C., Whelan, B., Sukkarieh, S., 2016. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. Comput. Electron. Agric. 130, 83–96. https://doi.org/10.1016/j.compag.2016.09.014

Vasconez, J.P., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2020. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. Comput. Electron. Agric. 173, 105348. https://doi.org/10.1016/j.compag.2020.105348

Villacrés, J.F., Auat Cheein F., 2020. Detection and characterization of cherries: A deep learning usability case study in Chile. Agronomy 10(6), 835. https://doi.org/10.3390/agronomy10060835

Villacrés, J., Viscaino, M., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2023. Apple orchard production estimation using deep learning strategies: A comparison of tracking-by-detection algorithms. Comput. Electron. Agric. 204, 107513. https://doi.org/10.1016/j.compag.2022.107513

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., 2018. Deep learning for computer vision: A brief review. Comput. Intell. Neurosci. 2018, 7068349. https://doi.org/10.1155/2018/7068349

Walsh, K.B., Blasco, J., Zude-Sasse, M., Sun X., 2020. Review: Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use. Postharvest Biol. Technol 168, 111246. https://doi.org/10.1016/j.postharvbio.2020.111246

Walsh, K., Payne, A., Subedi, P., Anderson, N., 2021. Mango fruit on tree image collection [WWW Document]. CQUniversity. https://doi.org/10.4226/145/59c9b2448242c

Wang, C., Lee, W.S., Zou, X., Choi, D., Gan, H., Diamond, J., 2018a. Detection and counting of immature green citrus fruit based on the Local Binary Patterns (LBP) feature using illumination-normalized images. Precis. Agric. 19(6), 1062–1083. https://doi.org/10.1007/s11119-018-9574-5

Wang, D., He, D., 2021. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. Biosyst. Eng. 210, 271-281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

Wang, D., He, D., 2022. Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background. Comput. Electron. Agric. 196, 106864. https://doi.org/10.1016/j.compag.2022.106864

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, H., Zheng, X., 2019. Survey of deep learning based object detection, in: 2nd International Conference on Big Data Technologies (ICBDT2019), pp. 149–153. https://doi.org/10.1145/3358528.3358574

Wang, Q., Nuske, S., Bergerman, M., Singh, S., 2013. Automated crop yield estimation for apple orchards, in: The 13th International Symposium on Experimental Robotics, pp. 745–758. https://doi.org/10.1007/978-3-319-00065-7_50

Wang, X., Tang, J., Whitty, M., 2021. DeepPhenology: Estimation of apple flower phenology distributions based on deep learning. Comput. Electron. Agric. 185, 106123. https://doi.org/https://doi.org/10.1016/j.compag.2021.106123

Wang, X., Tang, J., Whitty, M., 2022. Data-centric analysis of on-tree fruit detection: Experiments with deep learning. Comput. Electron. Agric. 194, 106748. https://doi.org/10.1016/j.compag.2022.106748

Wang, Z., Koirala, A., Walsh, K., Anderson, N., Verma, B., 2018b. In field fruit sizing using a smart phone application. Sensors, 18(10), 3331. https://doi.org/10.3390/S18103331

Wang, Z., Walsh, K., Koirala, A., 2019. Mango fruit load estimation using a video based MangoYOLO—Kalman filter—Hungarian algorithm method. Sensors 19(12), 2742. https://doi.org/10.3390/s19122742

Wang, Z., Walsh, K.B., Verma, B., 2017. On-tree mango fruit size estimation using RGB-D images. Sensors, 17(12), 2738. https://doi.org/10.3390/s17122738

Ward, I. R., Laga, H., Bennamoun, M., 2019. RGB-D Image-based object detection: From traditional methods to deep learning techniques, in: Rosin, P., Lai, Y.K., Shao, L., Liu, Y. (Eds.), RGB-D Image Analysis and Processing. Springer, Cham, pp. 169–201. https://doi.org/10.1007/978-3-030-28603-3_8

Wendel, A., Underwood, J., Walsh, K., 2018. Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. Comput. Electron. Agric. 155, 298–313. https://doi.org/10.1016/j.compag.2018.10.021

Williams, H.A.M., Jones, M.H., Nejati, M., Seabright, M.J., Bell, J., Penhall, N.D., Barnett, J.J., Duke, M.D., Scarfe, A.J., Seok, H., Lim, J., Macdonald, B.A., 2019. Robotic kiwifruit harvesting using machine vision , convolutional neural networks , and robotic arms. Biosyst. Eng. 181, 140–156. https://doi.org/10.1016/j.biosystemseng.2019.03.007

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric, in 2017 IEEE International Conference on Image Processing (ICIP), pp. 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962

Wu, G., Li, B., Zhu, Q., Huang, M., Guo, Y., 2020. Using color and 3D geometry features to segment fruit point cloud and improve fruit recognition accuracys. Comput. Electron. Agric. 174, 105475. https://doi.org/10.1016/j.compag.2020.105475

Xie, D., Chen, L., Liu, L., Chen, L., Wang, H., 2022. Actuators and sensors for application in agricultural robots: A review. Machines 10(19), 913. https://doi.org/10.3390/machines10100913

Yan, B., Fan, P., Lei, X., Liu, Z., Yang, F., 2021. A real-time apple targets detection method for picking robot based on improved YOLOv5. Remote Sens. 13(9), 1619. https://doi.org/10.3390/rs13091619

Yeo, S.T., Edwards, R.T., 2006. Encouraging fruit consumption in primary schoolchildren: a pilot study in North Wales, UK. J. Hum. Nutr. Diet. 19, 299-302. https://doi.org/10.1111/j.1365-277X.2006.00706.x

Yu, L., Xiong, J., Fang, X., Yang, Z., Chen, Y., Lin, X., Chen, S., 2021. A litchi fruit recognition method in a natural environment using RGB-D images. Biosyst. Eng. 204, 50–63. https://doi.org/10.1016/j.biosystemseng.2021.01.015

Yu, Q., Yang, H., Gao, Y., Ma, X., Chen, G., Wang, X., 2022a. LFPNet: Lightweight network on real point sets for fruit classification and segmentation. Comput. Electron. Agric. 194, 106691. https://doi.org/10.1016/j.compag.2022.106691

Yu, T., Hu, C., Xie, Y., Liu, J., Li, P., 2022b. 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

Yu, Y., Zhang, K., Yang, L., Zhang, D., 2019. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. Comput. Electron. Agric. 163, 104846. https://doi.org/10.1016/j.compag.2019.06.001

Zhang, C., Zou, K., Pan, Y., 2020. A method of apple image segmentation based on color-texture fusion feature and machine learning. Agronomy, 10(7), 972. https://doi.org/10.3390/agronomy10070972

Zhang, W., Chen, K., Wang, J., Shi, Y., Guo, W., 2021a. Easy domain adaptation method for filling the species gap in deep learning-based fruit detection. Hortic. Res. 8, 119.

Zhang, W., Liu, Y., Chen, K., Li, H., Duan, Y., Wu, W., Shi, Y., Guo, W., 2021b. Lightweight fruit-detection algorithm for edge computing applications. Front. Plant Sci. 12, 740936. https://doi.org/10.3389/fpls.2021.740936

Zhang, Y., Yu, J., Chen, Y., Yang, W., Zhang, W., He, Y., 2022a. Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application. Comput. Electron. Agric. 192, 106586. https://doi.org/10.1016/j.compag.2021.106586

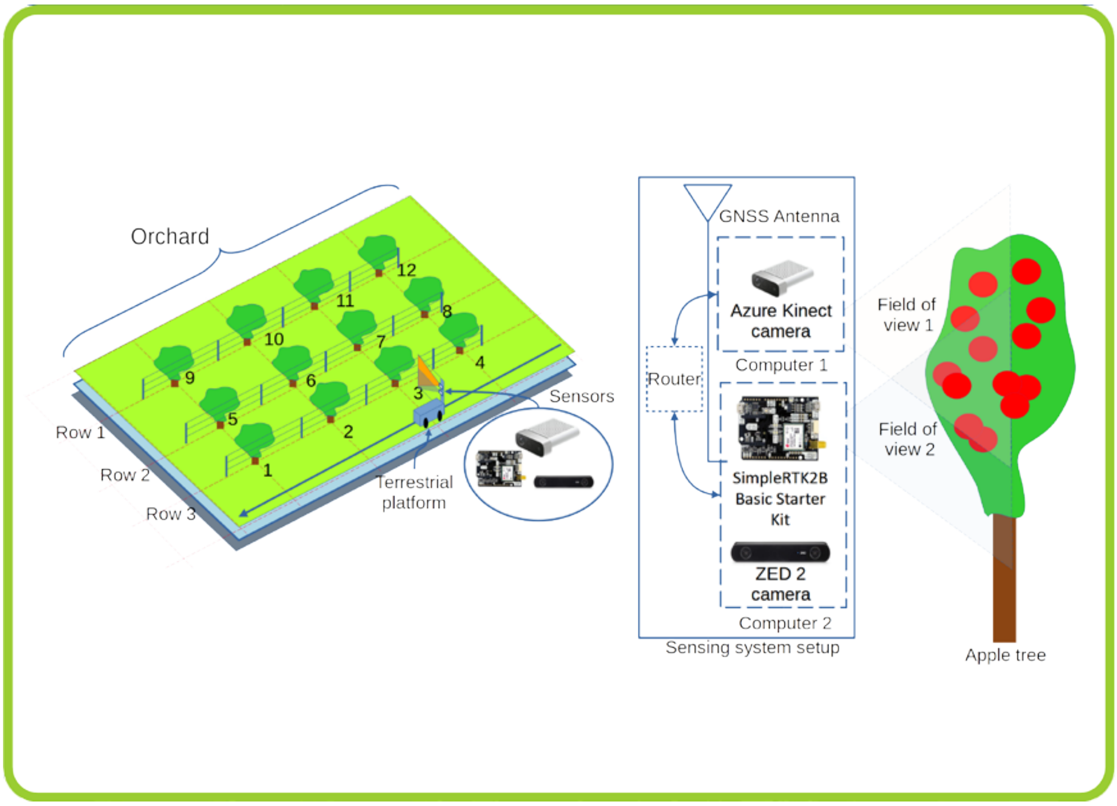
Zhang, Y., Zhang, W., Yu, J., He, L., Chen, J., He, Y., 2022b. Complete and accurate holly fruits counting using YOLOX object detection. Comput. Electron. Agric. 198, 107062. https://doi.org/10.1016/j.compag.2022.107062

Zhao, Y., Gong, L., Huang, Y., Liu, C., 2016a. A review of key techniques of vision-based control for harvesting robot. Comput. Electron. Agric. 127, 311–323. https://doi.org/10.1016/j.compag.2016.06.022

Zheng, C., Abd-Elrahman, A., Whitaker, V., 2021. Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming. Remote Sens. 13(3), 531. https://doi.org/10.3390/rs13030531

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M., 2020. Graph neural networks: A review of methods and applications. AI Open 1, 57-81. https://doi.org/10.1016/j.aiopen.2021.01.001

Chapter 4: AKFruitData: a dual software application for Azure Kinect cameras to acquire and extract informative data in yield tests performed in fruit orchard environments



This chapter was published as:
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

AKFruitData: a dual software application for Azure Kinect cameras to acquire and extract informative data in yield tests performed in fruit orchard environments

Juan Carlos Miranda*, Jordi Gené-Mola, Jaume Arnó, Eduard Gregorio

Research Group in AgroICT & Precision Agriculture, Department of Agricultural and Forest Engineering, Universitat de Lleida (UdL) - Agrotecnio CERCA Center, Rovira Roure 191, 25198 Lleida, Catalonia, Spain

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

Abstract

The emergence of low-cost 3D sensors, and particularly RGB-D cameras, together with recent advances in artificial intelligence, is currently driving the development of in-field methods for fruit detection, size measurement and yield estimation. However, as the performance of these methods depends on the availability of quality fruit datasets, the development of ad-hoc software to use RGB-D cameras in agricultural environments is essential. The AKFruitData software introduced in this work aims to facilitate use of the Azure Kinect RGB-D camera for testing in field trials. This software presents a dual structure that addresses both the data acquisition and the data creation stages. The acquisition software (AK_ACQS) allows different sensors to be activated simultaneously in addition to the Azure Kinect. Then, the extraction software (AK_FRAEX) allows videos generated with the Azure Kinect camera to be processed to create the datasets, making available colour, depth, IR and point cloud metadata. AKFruitData has been used by the authors to acquire and extract data from apple fruit trees for subsequent fruit yield estimation. Moreover, this software can also be applied to many other areas in the framework of precision agriculture, thus making it a very useful tool for all researchers working in fruit growing.

Keywords: RGB-D camera, data acquisition, data extraction, fruit yield trials, precision fructiculture.

Metadata

Nr	Code metadata	
C1	Current code version	1.0
C2	Permanent link to code/repository used for this code version	https://github.com/GRAP-UdL-AT/ak_acquisition_system https://github.com/GRAP-UdL-AT/ak_frame_extractor
C3	Permanent link to reproducible capsule	NA
C4	Legal code license	MIT license (http://opensource.org/licenses/MIT)
C5	Code versioning system used	Github
C6	Software code languages, tools and services used	Python 3.8 or later Required packages managed with pip: requeriments_win.txt, requirements_linux.txt Azure Kinect SDK, Stereolab ZED SDK, Pyk4a
C7	Compilation requirements, operating environments and dependencies	Windows 10 or Ubuntu Linux 20.04, Azure Kinect SDK, Python 3.8 or later.
C8	If available, link to developer documentation/manual	https://github.com/GRAP-UdL-AT/ak_acquisition_system/blob/main/README.md https://github.com/GRAP-UdL-AT/ak_frame_extractor/blob/main/README.md
C9	Support email for questions	juancarlos.miranda@udl.cat

4.1 Motivation and significance

Over the coming years, a significant food production increase will be required to meet the demand of a world population that could reach 9.7 billion people by mid-century [1]. To achieve this, environmentally friendly techniques must be applied, being at once socially and economically sustainable [2]. Advances in agricultural robotics [3], artificial intelligence [4], sensor integration [5], and big data [6], together with the application of precision farming techniques [7,8] promise to be essential to overcome these challenges.

In the fruit growing domain, the need for productivity improvement and sustainable resource management is currently driving the development of automatic methods for fruit detection and sizing, as well as for yield estimation [9–11]. These tasks are commonly performed by visual counting or manual measuring in certain trees sampled within the plot. The labour requirements are a clear weakness of this process, as well as the error (bias) that tends to be associated to what is a basically repetitive and subjective task under field conditions [12,13].

To test the usability of depth cameras (particularly RGB-D cameras) for the above purposes of detecting, measuring and quantifying fruit yield in apple trees, the Precision Agriculture research group of the University of Lleida (Catalonia, Spain) has launched a specific task on this topic within the PAgFRUIT national project. More specifically, the Azure Kinect camera (Microsoft, Redmond, WA) has been assessed to, in a first stage, acquire data on apples and, in a second stage, to extract information on their geometry and create a database for later analysis. Hence the dual structure of two different software tools, while also making use of deep learning algorithms in combination with a global navigation satellite system (GNSS) so that data captured during the field test can be georeferenced. Fig. 4.1a shows a diagram of the assembly mounted on a terrestrial platform that can be moved along the rows.

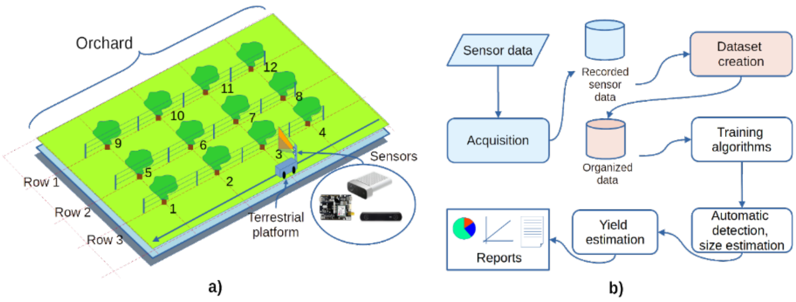


Fig. 4.1. Fruit yield estimation using computer vision methods. a) Sensor assembly scheme in field tests. b) Proposed stages of data acquisition and extraction.

Fig. 4.1b shows the two stages mentioned above, i) data acquisition and, ii) dataset creation. The developed software therefore focuses on these two stages, which are considered previous and essential to facilitate the use of RGB-D cameras for testing in field trials. With a complete database made available, the user can then train and validate artificial intelligence algorithms for detection, size measurement and yield estimation in apple trees. More specifically, the software for the acquisition stage (AK_ACQS) allows different sensors, such as time-of-flight (ToF) and stereo vision cameras, combined with a GNSS receiver to be activated simultaneously. The second extraction software (AK_FRAEX) allows videos generated with the ToF camera to be processed to create the aforementioned datasets. The selected ToF camera is the Azure Kinect DK, currently one of the most reputable RGB-D sensors used for commercial projects [14] and outdoor agricultural applications [15]. Color images, IR images, depth data and IMU motion data are examples of the multimodal information supplied by this camera. The ZED 2 camera (Stereolabs, San Francisco, CA) and the SimpleRTK2B Basic Starter Kit (Ardusimple, Lleida, Spain) are, respectively, the stereo vision camera and the GNSS receiver used in this work.

With respect to the first software (AK_ACQS), challenges posed by the PAgFRUIT project include acquisition from different sensors with different technologies, the georeferencing of data and the synchronization of the entire process. This has meant additional challenges such as: i) manufacturer software development kits that use different operating systems and programming languages, ii) sensors that require high processing demands, iii) sensors that must be used separately for reasons of hardware compatibility, and iv) sensors that require minimal hardware for operation and data recording. Fig. 4.2 helps to understand the difficulties involved in this multilevel acquisition using two different sensors together with the additional signal from the satellite receiver. The appropriate selection of libraries, recording formats and programming languages is especially important to meet the above requirements, aiming to provide a robust and easy-to-use solution even when real-time acquisition is performed by moving the ground platform during the acquisition stage.

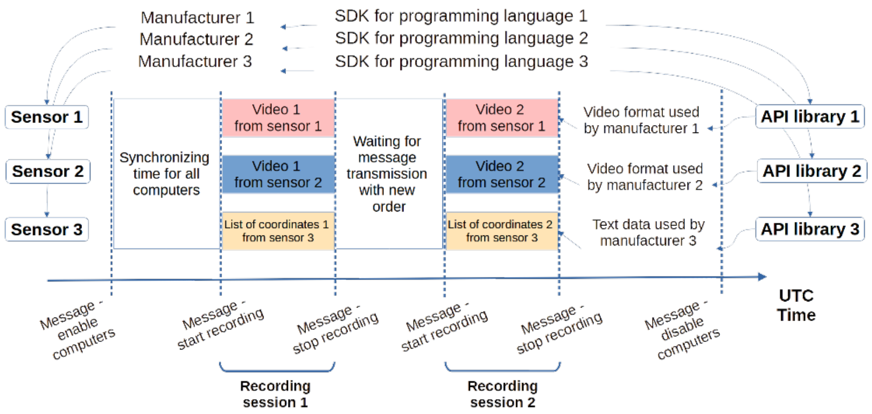


Fig. 4.2. Challenges posed by the use and synchronization of different sensors from different manufacturers and using different technologies in the first stage of fruit data acquisition.

As for the second software (AK_FRAEX), the user will have access to video frame information that can be extracted and stored in different directories, making available metadata involving colour, depth, IR and point clouds of the scenes. The generated metadata is labelled using external tools for their subsequent transformation by the software, at the request of the user, to train neural networks for yield detection and estimation. More specifically, conversion is from comma-separated values (.CSV) to extensible mark-up language (.XML) formats.

4.2 Software description

The tools developed under the PAgFRUIT research project are AK_ACQS and AK_FRAEX. The programming language used is Python 3.8 with GUIs based on Tkinter, which allows cross-platform execution on Windows 10 or later, and Linux operating systems.

4.2.1 Description of the AK_ACQS software

AK_ACQS is a software solution for data acquisition in fruit orchards using a sensor system mounted on a terrestrial vehicle (Fig. 4.3a). It allows the coordination of computers and sensors through the sending of remote commands via a GUI. At the same time, it adds an abstraction layer to the library stack of each sensor, facilitating its integration.

This software solution is supported by a local area network (LAN), which connects computers and sensors from different manufacturers (cameras of different technologies, GNSS receiver) for in-field fruit yield testing.

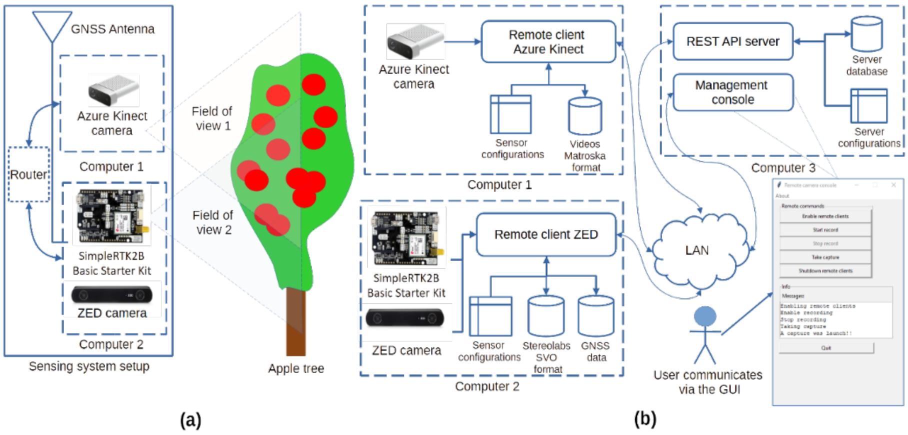


Fig. 4.3. Setup for data acquisition: (a) sensor system that is boarded on a terrestrial vehicle; (b) AK_ACQS software architecture showing the connectivity between the system components.

4.2.1.1 AK_ACQS software architecture

AK_ACQS is based on a client-server architecture (Fig. 4.3b) and uses the representational state transfer (REST) style to exchange messages over the HTTP protocol [16]. It is made up of a REST API server, remote clients and a management console.

The REST API server acts as an intermediary in the management of messages between remote clients and the management console, and stores information about the status of the instructions sent and received in the SQLite database [17]. The services are consumed by other components through APIs in JSON format [18] (Fig. 4.4a), assembled with the Django REST framework [19].

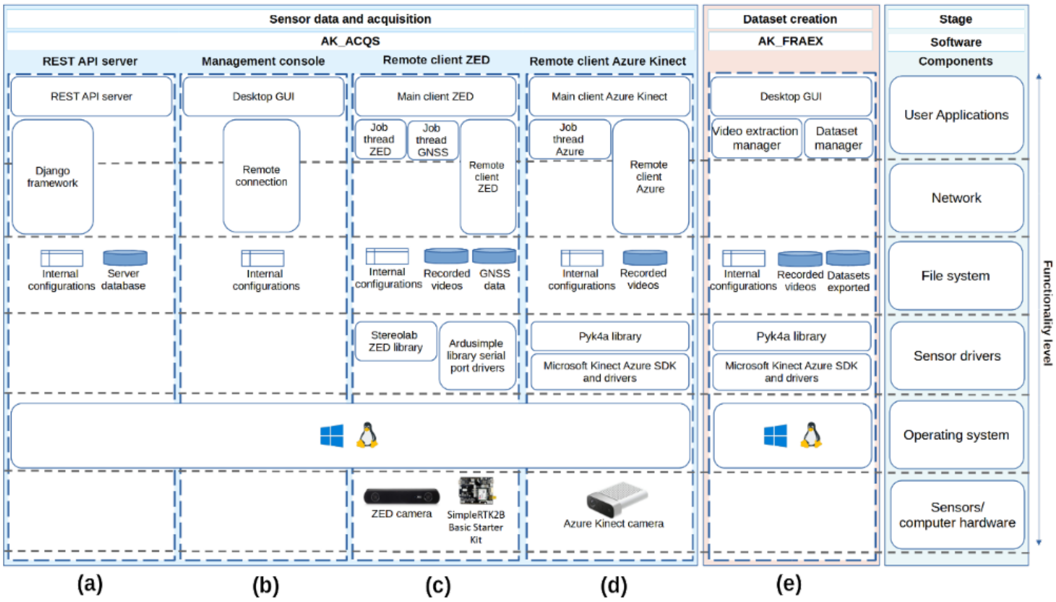


Fig. 4.4. Software components of the data acquisition and dataset creation tools. (a) REST API central server. (b) Management console. (c) Remote client ZED camera. (d) Remote client Azure Kinect camera. (e) AK_FRAEX data extraction tool.

The management console (Fig. 4.4b) offers the user the possibility of sending instructions to the devices and displaying the system status on the screen through the “Desktop GUI” package. It is networked with the server to send and receive messages, making use of the “Remote connection” class. The initial settings are read by the console from a text file. Remote clients listen to instructions from the server and interpret the commands to be applied to a given sensor. They are responsible for storing the recording sessions, configuring the devices and communicating with the hardware through the manufacturer's libraries.

The architecture of the remote clients (Fig. 4.4c, d) consists of the components "Remote client ZED" and "Remote client Azure Kinect". The processes are launched by "Main client ZED" and "Main client Azure" respectively, invoking the classes "Remote client ZED" and "Remote client Azure", which are containers of the function loop to make calls to the REST APIs. These classes interpret the commands and transfer them to the low-level libraries, and are also responsible for managing the stored configurations (connection data, registration parameters for each type of sensor). Communication with the sensors is done through the "Job thread ZED", "Job thread GNSS" and "Job thread Azure" classes, which group calls to the SDK functions of each manufacturer and launch parallel activities for data collection.

At the lowest level are the libraries of the manufacturer of each sensor and the specific drivers that access the hardware. In the case of the ZED 2 camera [20] and the Ardusimple SimpleRTK2B – Basic Starter Kit receiver [21], the programming kits are offered with development support for the Python language in the operating systems Ubuntu Linux 18/20, Windows 10/11 and NVIDIA JetPack. In the case of the Azure Kinect camera [22], initial support is oriented to C/C++ languages for Ubuntu Linux 18/20 and Windows. In this work, the third-party library called pyk4a [23] has been used, which is a Python wrapper over the original functions of the SDK.

Each manufacturer records data from their devices under a different format. The ZED 2 camera uses the proprietary format called SVO [24]. Videos produced with the Azure Kinect DK camera record data in Matroska (MKV) format [25], and GNSS receiver data is saved in text files.

Under the aforementioned architecture, the components can be hosted on different computers. Each computer connected to the network synchronizes its internal clock to Universal Time Coordinated (UTC) with Network Time Protocol (NTP) [26]. The time values of the synchronized clocks are used in the post-processing of the videos captured in the field.

The ACQS software architecture has been designed to be able to add other types of devices. Therefore, a generic remote client (remote_client_generic/) is included in the source code, which can be programmatically extended to support other types of devices or cameras. This option is beyond the scope of this work, but is left open for future applications.

4.2.1.2 AK_ACQS software functionalities

The functionalities of AK_ACQS consist of remotely enabling and disabling clients, taking snapshots and starting and stopping video recordings, as well as logging latitude and longitude coordinates during the video recording time.

- “ENABLE REMOTE CLIENTS” allows the user to send an attention call to the devices to configure them in the initial state of listening to orders.
- “TAKE CAPTURES” makes it easy for the user to capture short videos, automatically starting and stopping video or snapshot recording.

- “START RECORDING”/“STOP RECORDING” is the functionality that allows to send start and stop recording messages to all connected clients. Remote clients managing a GNSS receiver will start/stop operations for coordinated capture.
- "DISABLE REMOTE CLIENTS" allows user to remotely turn off the devices. These will stop operating when receiving and processing the message.

The recorded files are stored on the host computers, just like the data collected by the GNSS receiver.

4.2.2 Description of the AK_FRAEX software

AK_FRAEX is a desktop tool created for post-processing tasks after field acquisition. It enables the extraction of information from videos recorded in MKV format with the Azure Kinect camera. Through a GUI, the user can configure initial parameters to extract frames and automatically create the necessary metadata for a set of images.

4.2.2.1 AK_FRAEX software architecture

The AK_FRAEX tool (Fig. 4.4e) presents the functionalities of the application to the user via the "Desktop GUI Tkinter" package. Here, the functions are grouped together to offer the user the possibility to easily extract data, which would otherwise have to be done programmatically by calling functions from the manufacturer's SDK API.

The GUI makes use of the methods programmed in the "Video extraction manager" package, which contains primitives that are transferred to the pyk4a base library to access the MKV-type files. The functions included in this package are: obtaining information about the video, exporting frames to files with and without size transformations, exporting frames to point cloud files, helpers for colour format conversion.

The tasks that involve access to the file system and directory management are grouped in the "Dataset manager". The functions included in this package are: creation of the dataset hierarchy, parsing of file formats and data migration.

4.2.2.2 AK_FRAEX software functionalities

In AK_FRAEX it is possible to select one or several video files (batch of files) and extract RGB, depth, IR images, and cloud point data. The datasets created with this tool can be used in image analysis and/or object detection tasks.

The software functionalities were grouped into three main tabs, in order to intuitively guide users through the extraction process.

- "DATA SET CREATION" allows the user to create a new metadata hierarchy. This consists of subfolders that will be filled with data extracted from the videos. Providing the user with an organised structure for storing RGB images, depth data, IR images, segmentation masks and annotation files in CSV and XML format of the PASCAL-VOC type [27].
- "DATA EXTRACTION" offers the user the possibility to select a working folder and extract data from a batch of files, or from a specific video. The user can configure the starting frame and the number of frames to be exported by entering values in the data fields. It is also possible to extract cloud points by selecting a check button. This functionality can only be enabled for videos captured under the BGRA recording mode of the Azure Kinect camera.
- "DATA MIGRATION" allows the user to select a folder with files in CSV format and convert them into XML files under PASCAL-VOC format. This functionality was intended to be used in conjunction with the Pychet Labeller tool [28], to convert the data labelled by this software to XML format.

4.3 Illustrative examples

4.3.1 AK_ACQS: capturing data from fruit orchards

The configuration of this example consists of a ToF camera (Azure Kinect DK), a stereo vision camera (ZED 2) and a GNSS receiver (SimpleRTK2B – Basic Starter Kit). The devices are connected to two computers via a USB connection. Computer 1 is a Modern 15 A10RBS-484XES (MSI, New Taipei, Taiwan) laptop running Windows 10 and is used to host the management console and Azure Kinect camera functions. Computer 2 is a Jetson Xavier NX (NVIDIA, Santa Clara, CA) embedded computer using Ubuntu Jetpack and its purpose is to manage the GNSS receiver, the ZED 2 camera and host

REST API central server. Both computers receive IP addresses through Dynamic Host Configuration Protocol (DHCP) [29] and synchronize their clocks to UTC servers. The aforementioned elements come together in a vehicle-mounted Wi-Fi LAN network, where a Redmi Note 8T (Xiaomi, Beijing, China) mobile phone performs the functions of router and DHCP and enables requests to NTP services.

Assuming that the software components (REST API server, remote management console, remote clients) are in the running state and with network visibility, it will be necessary to establish the IP address of the server so that the components can now direct their requests to this central server remaining operational. The configuration used in this example can be seen in Fig. 4.5, which also shows the management console and the hardware used.

Once the software and hardware requirements are ready, the user can start the operations from the GUI of the management console. The first step is to enable the remote clients so that they remain listening for new instructions, using the "ENABLE REMOTE CLIENTS" button. From this moment onwards, the user can choose to take short video captures with the "TAKE CAPTURE" button, or proceed to make recordings using the "START RECORD" and "STOP RECORD" buttons. Finally, all operations can be stopped with the "DISABLE REMOTE CLIENTS" button, a function that disconnects the remote clients from all the computers.

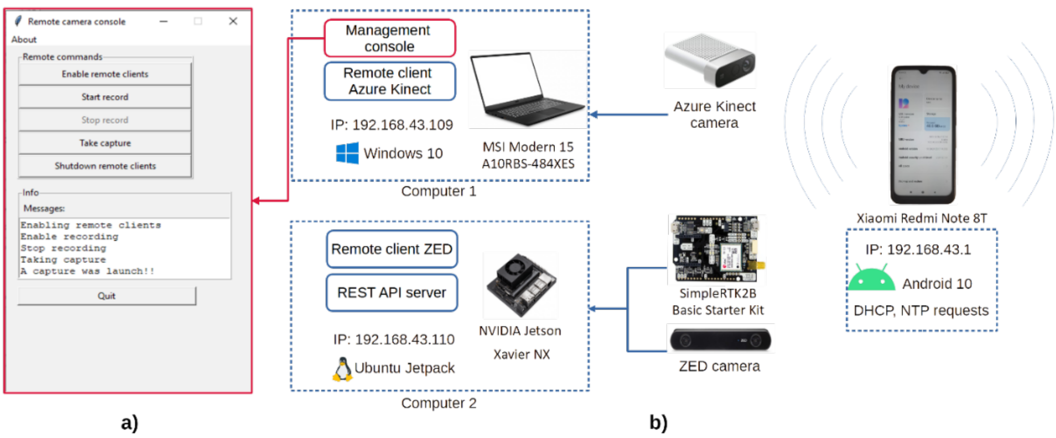


Fig. 4.5. Capturing fruit data using the AK_ACQS software. a) Console manager screen. b) Hardware setup elements used in field trials.

4.3.2 AK_FRAEX: creating datasets for fruit yield analysis

Using other tools mentioned below, it is possible to build datasets with labelled objects for neural network training processes. An example of this can be seen in Fig. 4.6, in which AK_FRAEX (information extraction) interacts with Pychet Labeller (object labelling), MATLAB Image Segmenter (mask creation) [30] and CloudCompare (cloud points processing) [31] through a series of steps.

First, a frame extraction from videos in MKV format (Fig. 4.6a) is applied to automatically create a base directory hierarchy with the exported data (Fig. 4.6b). Then, the objects are labelled using the Pychet Labeller tool, the data of which is stored in CSV format (Fig. 4.6c). Binary masks are then created in portable network graphics (PNG) format (Fig. 4.6d). Finally, the conversion of the generated metadata to XML format is applied, thus leaving a database ready with objects and their references for further analysis (Fig. 6e). The data can be read by scripts of different programming languages (Fig. 4.6f) and the extracted cloud point of each frame can be visualized (Fig. 4.6g).

Source codes in MATLAB, R and Python languages are provided as supplementary material in the AK_FRAEX software to further manage the exported data using different procedures. Test videos are also attached to serve as tutorial examples for users of the software.

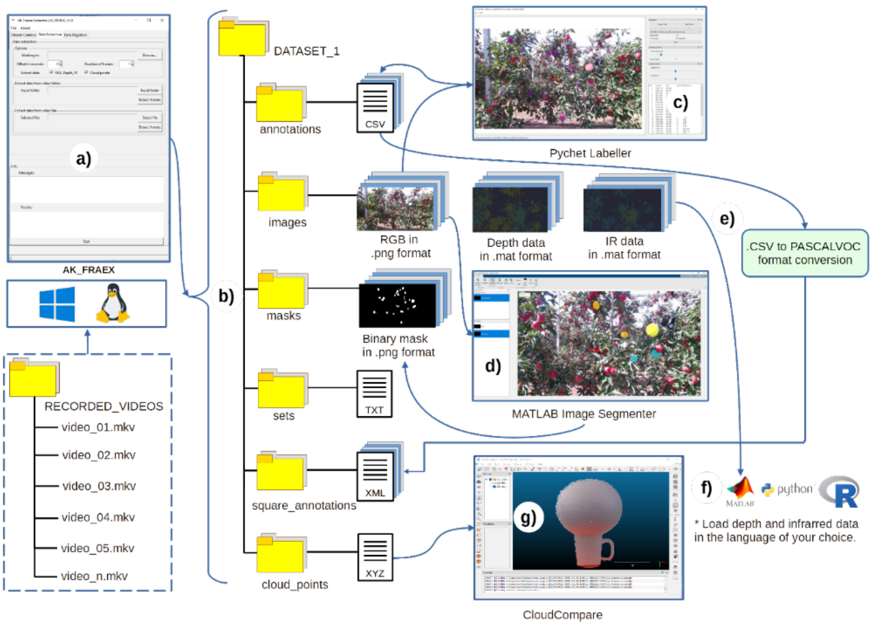


Fig. 4.6. Example of the functionality of the AK_FRAEX software together with external tools: (a) video data extraction, (b) automatic creation of directory hierarchies, (c) object labelling, (d) creation of binary masks, (e) conversion to XML format, (f) export for use with other programming languages, (g) visualization of cloud points of the scenes.

4.4 Impact

The impact of the AKFruitData software is expected to be important in agriculture as a tool that allows different optical sensors to be tested and validated for agronomic use in fruit orchards. Two basic research domains can take advantage of the use of ToF cameras and GNSS receivers through AKFruitData, namely i) for total fruit load estimation to adjust subsequent thinning operations, and ii) for fruit yield estimation just before harvest. Moreover, AKFruitData can be applied in multiple precision agriculture domains including, among others, leaf canopy phenotyping, fruit tree disease detection, branch detection or tree growth monitoring.

Primarily intended for fruit yield forecasting with red apple cultivars, AKFruitData could also serve as data acquisition and extraction software for other fruit crop species. Fruit growers and advisors are aware of the importance of accurately estimating fruit yield, with this currently being a key task to ensure the proper logistical arrangements of a harvesting process that usually covers a limited time window.

While aiming at a medium-long term economic impact associated to better farm management, AKFruitData is expected to become a highly useful tool for researchers working in fruit growing in the short term. Ultimately, AKFruitData aims to bring detailed digital information about fruit tree canopies, serving as a bridge between researchers, application developers and fruit growers. Through the use of this software, different data sources are made available so that, with suitable formats, different methods of analysis and the simulation of trees, fruits, sizes and distribution can be subsequently applied. With the development of this possibility of applying optical sensing in fruit growing, progress is expected and a positive impact is anticipated.

4.5 Conclusions

AKFruitData software has been developed for the primary use of Azure Kinect cameras in outdoor agricultural environments such as fruit orchards. Based on a modular concept, the software includes a first AK_ACQS module for data acquisition and a second AK_FRAEX module to extract information from videos recorded in MKV format. The main challenges overcome through the software design involve the simultaneous use and synchronization of different cameras and sensors, and the need to georeference data when acquisition is performed on-the-go within the plot to obtain mappable spatial information. The Python programming language has been used with Tkinter GUI, making it possible to run the software on Windows 10 and GNU/Linux operating systems. Although future improvements are planned under a concept of continuous updating of functionalities, the software presented provides RGB, depth, IR and point cloud data to test computer vision methods for efficient fruit detection, allowing estimation of fruit load and productivity. Other applications, such as leaf canopy phenotyping to optimize many other agricultural tasks, could also be addressed. In short, AKFruitData is presented as open-source software, with an essential focus on facilitating the use of the Azure Kinect cameras in agricultural research.

4.6 Declaration of competing interest

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

4.7 Acknowledgements

This work was funded by the Spanish Ministry of Science and Innovation and the Spanish Estate Research Agency through grant RTI2018-094222-B-I00 (PAgFRUIT project). This work was also supported by the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya under Grant 2017-SGR-646. The Secretariat of Universities and Research of the Department of Business and Knowledge of the Generalitat de Catalunya and Fons Social Europeu (FSE) are also thanked 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. The authors would also like to thank the Institut de Recerca i Tecnologia Agroalimentàries (IRTA) for allowing the use of their experimental fields, and in particular Dr. Luís Asín and Dr. Jaume Lordán who have contributed to the success of this work.

4.8 References

[1] U. Nations (2019). World Population Prospects 2019: Highlights, https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf [accessed Sep 05 2022].

[2] FAO (2017). The future of food and agriculture -Trends and challenges. http://www.fao.org/3/a-i6583e.pdf [accessed Sep 05 2022].

[3] Oliveira L.F.P., Moreira A.P., Silva M.F. (2021). Advances in Agriculture Robotics: A State-of-the-Art Review and Challenges Ahead. Robotics, https://doi.org/10.3390/robotics10020052.

[4] Sozzi M., Cantalamessa S., Cogato A., Kayad A., Marinello F. (2022). Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. Agronomy, https://doi.org/10.3390/agronomy12020319.

[5] Bazzi C.L., Jasse E.P., Graziano Magalhães P.S., Michelon G.K., de Souza E.G., Schenatto K., Sobjak R. (2019). AgDataBox API – Integration of data and software in precision agriculture. SoftwareX, https://doi.org/10.1016/j.softx.2019.100327.

[6] Sishodia R.P., Ray R.L., Singh S.K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sensing, https://doi.org/10.3390/rs12193136.

[7] Gebbers R., Adamchuk V.I. (2010). Precision Agriculture and Food Security. Science, http://doi.org/10.1126/science.1183899.

[8] Kerry R., Escolà A., eds. (2021). Sensing Approaches for Precision Agriculture. Springer Cham, https://doi.org/10.1007/978-3-030-78431-7.

[9] Koirala A., Walsh K.B., Wang Z., McCarthy C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric., https://doi.org/10.1016/j.compag.2019.04.017.

[10] Gené-Mola J., Sanz-Cortiella R., Rosell-Polo J.R., Morros J.-R., Ruiz-Hidalgo J., Vilaplana V., Gregorio E. (2020). Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry. Comput. Electron. Agric., https://doi.org/10.1016/j.compag.2019.105165.

[11] 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., https://doi.org/10.1016/j.compag.2021.106343.

[12] Zhang Z., Heinemann P.H., Liu J., Baugher T.A., Schupp J.R. (2016). The Development of Mechanical Apple Harvesting Technology: A Review. Trans. ASABE, https://doi.org/10.13031/trans.59.11737.

[13] Zhao Y., Gong L., Huang Y., Liu C. (2016). A review of key techniques of vision-based control for harvesting robot. Comput. Electron. Agric., http://dx.doi.org/10.1016/j.compag.2016.06.022.

[14] Terven J.R, Córdova-Esparza D.M. (2021). KinZ an Azure Kinect toolkit for Python and Matlab. Sci. Comput. Program., https://doi.org/10.1016/j.scico.2021.102702.

[15] Neupane C, Koirala A, Wang Z, Walsh KB. Evaluation of Depth Cameras for Use in Fruit Localization and Sizing: Finding a Successor to Kinect v2. Agronomy 2021; 11(9): 1780. doi: 10.3390/agronomy11091780.

[16] Fielding R, Gettys J, Mogul J, Frystyk H, Masinter L, Leach P, Berners-Lee T. RFC 2616: Hypertext Transfer Protocol -- HTTP/1.1, 1999, https://doi.org/10.17487/RFC2616.

[17] SQLite Consortium. SQLite, https://www.sqlite.org/ [accessed Sep 05 2022].

[18] Bray E. RFC 8259: The JavaScript Object Notation (JSON) Data Interchange Format, 2017, https://doi.org/10.17487/RFC8259.

[19] Encode OSS Ltd. Django REST Framework, https://www.django-rest-framework.org/ [accessed Sep 05 2022].

[20] Stereolabs. ZED SDK, https://www.stereolabs.com/developers/release/ [accessed Sep 05 2022].

[21] Ardusimple. simpleRTK2B - Basic Starter Kit, https://www.ardusimple.com/product/simplertk2b-basic-starter-kit-ip65/ [accessed Sep 05 2022].

[22] Microsoft. Azure Kinect DK documentation, https://docs.microsoft.com/en-us/azure/kinect-dk/ [accessed Sep 05 2022].

[23] Asselin LP, Martin O, etiennedub, Gruzinov I, DominikMa, von Forstner J, Boulanger S, (2021). https://doi.org/10.5281/ZENODO.4774311.

[24] Stereolabs. How do I convert SVO files to AVI, or image/depth sequences?, https://support.stereolabs.com/hc/en-us/articles/360009986754-How-do-I-convert-SVO-files-to-AVI-or-image-depth-sequences- [accessed Sep 05 2022].

[25] The Matroska Association. Matroska, https://www.matroska.org/ [accessed Sep 05 2022].

[26] Mills D, Burbank J, Kasch W. RFC 5905: Network Time Protocol Version 4: Protocol and Algorithms Specification, 2010. https://doi.org/10.17487/rfc5905.

[27] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. Int. J. Comput. Vis., https://doi.org/10.1007/s11263-009-0275-4.

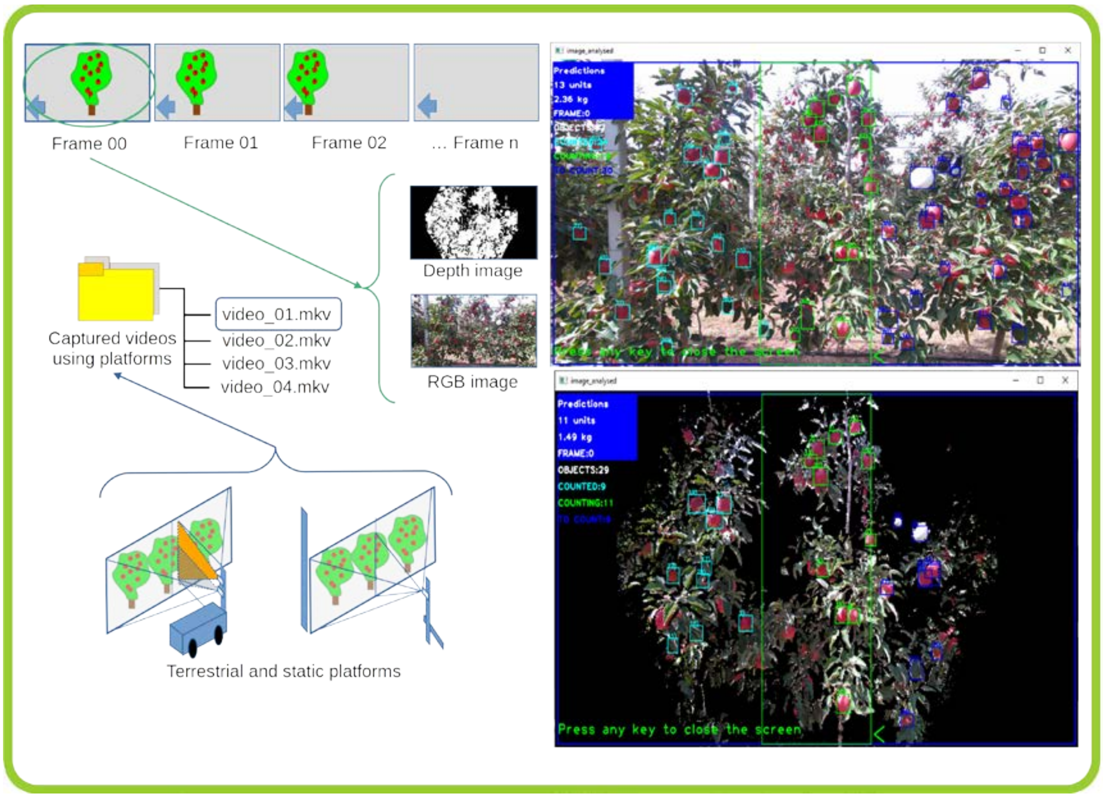
[28] Bargoti S., Pychet Labeller - An object annotation toolbox, https://github.com/acfr/pychetlabeller [accessed Sep 05 2022].

[29] Droms R. RFC 2131: Dynamic Host Configuration Protocol, 1997. https://doi.org/10.17487/rfc2131

[30] Mathworks. Image Segmenter, https://es.mathworks.com/help/images/ref/imagesegmenter-app.html [accessed Sep 05 2022].

[31] EDF R&D, CloudCompare, https://www.cloudcompare.org [accessed Sep 05 2022].

Chapter 5: AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards



This chapter was published as:
Miranda, J.C., Arnó, J., Gené-Mola, J., Fountas, S., Gregorio, E., 2023. AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards. SoftwareX 24, 101548. https://doi.org/https://doi.org/10.1016/j.softx.2023.101548

AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards

Juan Carlos Miranda[1,*], Jaume Arnó[1], Jordi Gené-Mola[1,2], Spyros Fountas[3], Eduard Gregorio[1]

[1] Research Group in AgroICT & Precision Agriculture, Department of Agricultural and Forest Sciences and Engineering, Universitat de Lleida (UdL) - Agrotecnio CERCA Center, Rovira Roure 191, 25198 Lleida, Catalonia, Spain

[2] Efficient Use of Water in Agriculture Program, Institute of AgriFood Research and Technology (IRTA), Fruitcentre, Parc Científic I Tecnològic Agroalimentari de Gardeny (PCiTAL), 25003 Lleida, Catalonia, Spain

[3] Agricultural University of Athens, Iera Odos 75, 118 55 Athina, Greece

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

Abstract

AKFruitYield is a modular software that allows orchard data from RGB-D Azure Kinect cameras to be processed for fruit size and fruit yield estimation. Specifically, two modules have been developed: i) AK_SW_BENCHMARKER that makes it possible to apply different sizing algorithms and allometric yield prediction models to manually labelled colour and depth tree images; and ii) AK_VIDEO_ANALYSER that analyses videos on which to automatically detect apples, estimate their size and predict yield at the plot or per hectare scale using the appropriate algorithms. Both modules have easy-to-use graphical interfaces and provide reports that can subsequently be used by other analysis tools.

Keywords: RGB-D camera, fruit detection, apple fruit sizing, yield prediction, allometry.

Metadata

Nr	Code metadata description	
C1	Current code version	1.0
C2	Permanent link to code/repository used for this code version	https://github.com/GRAP-UdL-AT/ak_sw_benchmarker https://github.com/GRAP-UdL-AT/ak_video_analyser
C3	Permanent link to reproducible capsule	https://pypi.org/project/ak-sw-benchmarker/ https://pypi.org/project/ak-video-analyser/
C4	Legal code license	MIT license (http://opensource.org/licenses/MIT)
C5	Code versioning system used	Github
C6	Software code languages, tools and services used	Python 3.8 or later Required packages managed with pip: requeriments_win.txt, requirements_linux.txt Azure Kinect SDK Pyk4a
C7	Compilation requirements, operating environments and dependencies	Windows 10 or Ubuntu Linux 20.04, Azure Kinect SDK, Python 3.8 or later
C8	If available, link to developer documentation/manual	https://github.com/GRAP-UdL-AT/ak_sw_benchmarker/blob/main/README.md https://github.com/GRAP-UdL-AT/ak_video_analyser/blob/main/README.md
C9	Support email for questions	juancarlos.miranda@udl.cat

5.1 Motivation and significance

In a recent article, Miranda et al. [1] developed AKFruitData, a dual software application available to users to facilitate the use of RGB-D (Azure Kinect) cameras in apple orchard environments. With this initial software, tree data acquisition (videos) was performed to allow a second phase of the creation of datasets for further analysis. Videos were recorded under field conditions containing information about colour, depth and IR data from the scene. In the attempt to give continuity to the previous application, this paper focuses on presenting a second software (AKFruitYield) that complements the previous one, now seeking a double research and agronomic objective: i) to allow the benchmarking of different apple fruit sizing algorithms and allometric models with the aim of providing recommendations for the future; and ii) to develop a video analysis tool for RGB-D cameras that, including fruit detection and sizing, allows for a reliable prediction of fruit yield in apple orchards (per hectare or within a plot).

In order to better organize logistics at the farm level, obtaining reliable early yield estimates is an objective that the use of these new technologies can help to achieve [2–4]. Even efforts have been carried out by publishing software packages in open access to better understand and manage the soil-crop-yield continuum within the framework of Precision Agriculture [5–7]. The use of photonics and computer vision in fruit growing to achieve the aforementioned objective is becoming increasingly popular [3]. The detection, measurement and estimation of fruit size are tasks that can be carried out using different methodologies that include those based on 2D images from those that provide 3D point clouds [8–11]. In the first case, it is necessary the use of calibration targets (known dimensions) that must be located close to the object (fruit) that is aimed to be sized. On the other hand, 3D techniques (e.g., structure-from-motion, LiDAR sensors, RGB-D cameras) generate 3D point cloud reconstructions of the scene. This avoids the use of calibration targets and allows simultaneous estimate the size of all the fruits in the image. Among 3D techniques, RGB-D cameras are a very interesting option given their low cost and the varied and extensive information they offer in each capture [12–15]. Computer techniques together with the use of RGB-D cameras have been applied to different crops [16–22] that are of interest both to the scientific community and industry. However, the pending issue is to have friendly computer applications (automated algorithms) to process the large amount of field data and convert it to useful information for fruit growers and managers. This is the goal of the AKFruitYield software presented in this paper.

The significance of the software that is presented lies in its addressing both fruit size estimation and allometric yield prediction following a comprehensive approach and allowing, at the same time, benchmarking of different algorithms in search of the best combined strategy (detection-sizing-yielding). Unlike certain commercial applications, the fact that information on estimation errors is provided at each stage is a very useful feature, especially for researchers and managers. Briefly, AKFruitYield software can be used as a research tool (increasing knowledge about automatic algorithms in fruit growing) or, from a more applied point of view, as a yield prediction tool covering a need that is increasingly in demand on the part of the fruit sector.

All the data used for the development of this software come from a field test carried out in an experimental apple orchard (cultivar Story® Inoredcov) located in Mollerussa, Lleida, Spain. Specifically, trees were arranged according to a 3.6 x 1 m plantation pattern with a canopy height of about 3.5 m (latitude: 41.617465 N; longitude: 0.870730; 246.3 m a.s.l. ETRS89). As mentioned, fruit data was acquired using the Azure Kinect camera (Microsoft, Redmond, WA, USA) on which the development of this software is focused.

5.2 Software description

The modular structure of AKFruitYield arises from a double requirement: i) to provide a tool for scientific use (AK_SW_BENCHMARKER), in this case a module to design and test different sizing and yield prediction algorithms for apple fruits; and ii) to complement the first with another second more applied tool (AK_VIDEO_ANALYSER) for automatic fruit detection (based on deep learning) on videos recorded with the Azure Kinect camera to then automate fruit sizing and yield prediction algorithms on the detected apples. Both modules (AK_SW_BENCHMARKER and AK_VIDEO_ANALYSER) were developed using the Python 3.8 programming language with Tkinter-based graphic user interfaces (GUIs), making it easier to use the applications on Windows 10 or later and Linux operating systems. In addition, image processing made use of the OpenCV library, with Numpy, Scikitlearn and Pandas being the libraries used for data management. With respect to the implementation of object detectors, PyTorch making use of Mask R-CNN [23] and Faster R-CNN

[24] models trained on own apple data from the experimental orchard was the open source learning library used in the video analyser module.

Fig. 5.1 shows the proposed data acquisition and extraction stages on which the design of AKFruitData and AKFruitYield was based. AKFruitData was designed exclusively for the acquisition and extraction of data from fruit orchards. Once the orchard data is obtained and conveniently organized (Fig. 5.1a), the AKFruitYield software can be put into operation to, after algorithm training, perform fruit detection, sizing and yield prediction. Interoperability between the specific modules of each software (AKFruitData and AKFruitYield) is shown in Fig. 5.1b. While the AK_ACQS module is responsible for data acquisition in video format from the orchard (complemented by the AK_SM_RECORDER [25]), the AK_FRAEX module allows extracting RGB images (frames) with additional depth information in different formats [1]. At this point the AKFruitYield software comes in, with the purpose of complementing the functionalities after the acquisition. Using the AK_SW_BENCHMARKER module on extracted frames (PASCAL-VOC format, [26]), different sizing algorithms and allometric models can be combined with final testing of results. In parallel, the AK_VIDEO_ANALYSER module implements the most appropriate sizing and yield prediction algorithms on video records (Matroska MKV format [27]), having previously trained fruit detection deep learning models on frames extracted in COCO format [28].

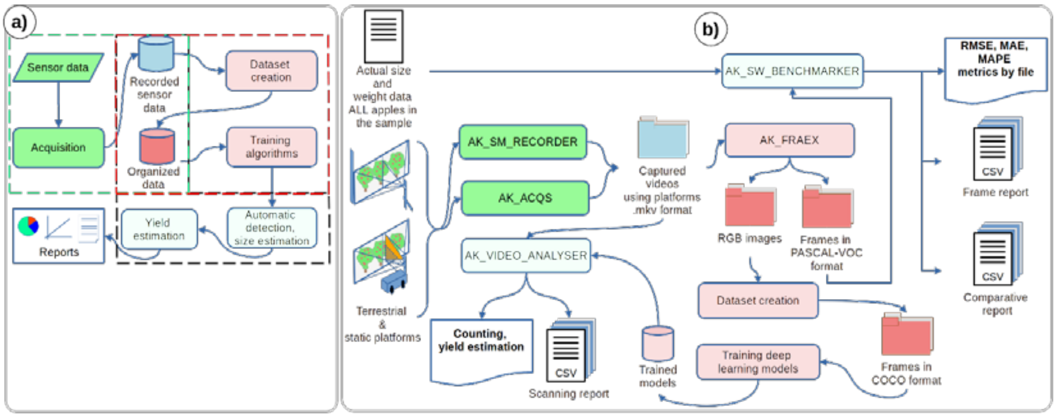


Fig. 5.1 a) Proposed stages of data acquisition and extraction for AKFruitData and AKFruitYield. Dashed green lines correspond to processes related to acquisition, red lines to processes related to data creation and training, and black lines to processes for performance estimation. b) Interoperability between the data acquisition (AK_ACQS; AK_SM_RECORDER), data creation (AK_FRAEX), algorithm benchmarking (AK_SW_BENCHMARKER) and video analysis (AK_VIDEO_ANALYSER) modules. The processes proposed in Figure 5.1 are expanded and represented by the developed software.

5.2.1 *AK_SW_BENCHMARKER module*

As mentioned, the AK_SW_BENCHMARKER module was designed for algorithm comparison tasks. Using the interface shown in Fig. 5.2, the software user can select between different size estimation algorithms that provide geometric measurements of the fruit (apple width and height). These size parameters are then used as inputs in different selectable allometric models for the final fruit yield prediction (weight in g).

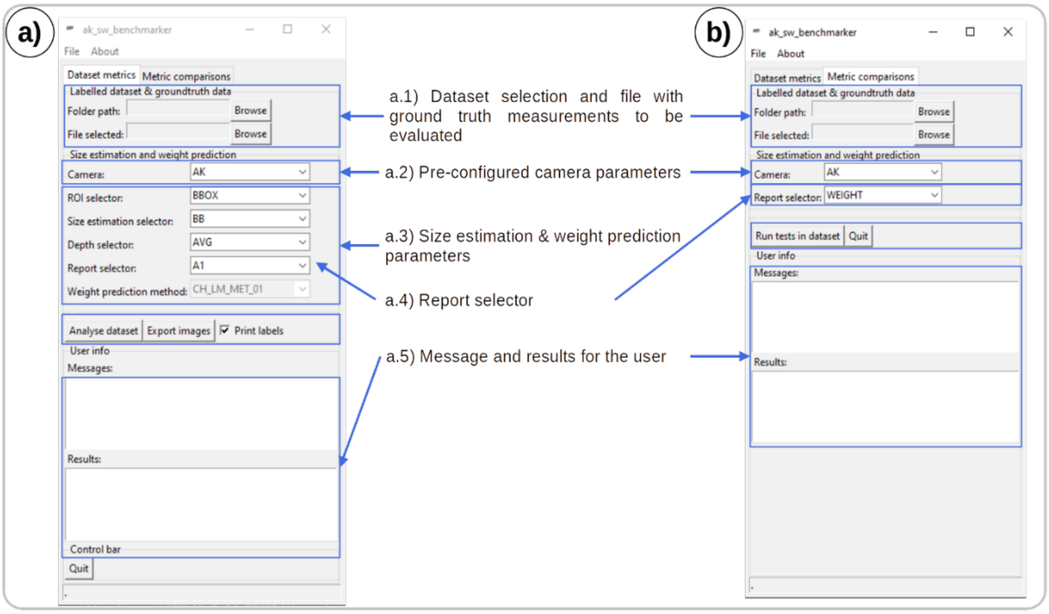


Fig. 5.2. AK_SW_BENCHMARKER module user interface. a) 'Dataset metrics' tab to select data (frames) and configure the sizing and yield prediction algorithms. b) 'Metric comparisons' tab to report results and error statistics.

The AK_SW_BENCHMARKER module offers the user two main tabs with grouped functionalities (Fig. 5.2): 'Dataset metrics' (Fig. 5.2a) and 'Metric comparisons' (Fig. 5.2b). Using the software starts with the selection of the dataset (labelled image in PASCAL-VOC format created by the previous AK_FRAEX module in the AKFruitData software), with the 'Dataset metrics' tab open (Fig. 5.2a). Once the camera parameters are pre-configured (Azure Kinect in our case), the user must establish the region of interest (ROI) selection method to be applied to the images. Two approaches are available ('ROI selector' in Fig. 5.2a): assigning bounding boxes to the detected fruits (BBOX method) or delimiting the fruit region by a binary mask (MASK method).

The next phase is to estimate the size in pixels of the perpendicular axes that define the geometry of the fruit (width or caliber and height). In the case of delimiting the ROI by means of the BBOX method, the caliber and height are obtained by adjusting the bounding boxes (BBs) to the labelled fruit. If, instead, a binary mask is used to delimit the ROI, the user can select between four different options ('Size estimation selector' in Fig. 5.2a) for adjusting geometric figures taking the ROI mask as a reference: circle enclosing (CE), circle fitting (CF), ellipse fitting (EF), and rotated rectangle (RR).

'Depth selector' (Fig. 5.2a) is the drop-down menu where the distance from the fruit to the RGB-D camera (depth) can be estimated by calculating a metric on the ROI pixels. Available metrics are the average depth (AVG), the modal depth (MOD) or the minimum depth (MIN). This data is then used to convert the geometric measurements of the fruit (pixels) to measurements of width and height in mm (thin lens theory). The final step is to choose an allometric model from those proposed in the drop-down menu below ('Weight prediction method' in Fig. 5.2a).

The functionalities developed in the AK_SW_BENCHMARKER module (Fig. 5.2) are listed below. In either case, summary information is presented on screen ('User info' section), and reports are stored as files in an output directory.

- "Analyse dataset" allows benchmarking to be performed with a final report in CSV format of size estimates and weight prediction. The user has the option of introducing a file with values of real dimensions of fruits (ground truth) to compare with the set of images that it is desired to be analysed. A final report with results (size and weight) grouped by image and fruit will be presented according to the selected parameters in addition to the following error metrics: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).
- "Export images" makes it possible to visualize the geometric fitting of the ROIs on the objects (fruits) to be measured. Outputs are colour images including binary masks of selected objects and fruit labeling. This functionality adds value to the software since the user can observe how sizing algorithms are applied to the images, enabling corrective adjustments in the algorithm configuration if necessary.
- "Run tests in dataset" calculates the test metrics. The user must first select the 'Metric comparisons' tab (Fig. 5.2b) and then, indicate the estimate to be analysed (Report selector in Fig. 5.2b), namely the major geometric axis of the fruit (A1), the minor axis (A2) or the weight

(WEIGHT). Since all sizing-yielding combinations are analysed, this functionality allows the method with least error to be determined, also obtaining a final ranking of sizing algorithms or ranking of sizing-allometric model combinations.

5.2.2 AK_VIDEO_ANALYSER module

The AK_VIDEO_ANALYSER software module is focused on analysing RGB-D Azure Kinect camera videos. Automatic detection of fruits, sizing of detected fruits and yield prediction (fruit weight) based on allometry are the tasks it performs. Under the same concept as the benchmarking module, a GUI (Fig. 5.3a) facilitates configuration of the parameters, allowing: i) setting the start and end of the video fragments; ii) filtering by zones and depth; iii) choice of object detector; and iv) choice of sizing and yield prediction methods. In real time, information about counted fruits and yield is displayed to the user as the frames of the video fragments are analysed (Fig. 5.3b).

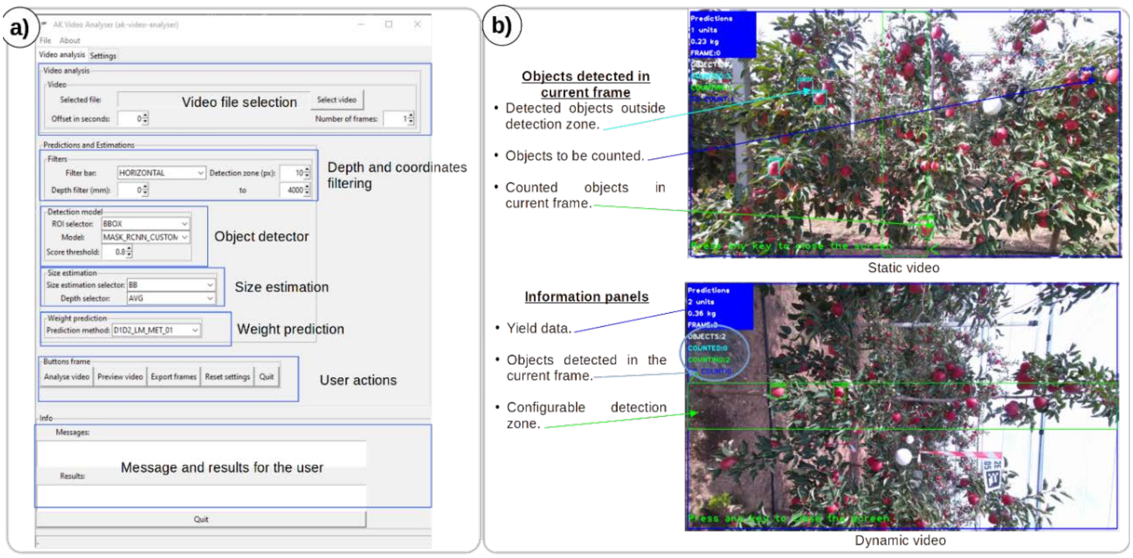


Fig. 5.3. AK_VIDEO_ANALYSER module user interface. a) Main GUI. b) Output screen showing detected fruits and report of results in real time.

Computerized tasks in the AK_VIDEO_ANALYSER module for processing information from video records to final yield results are shown in greater detail in Fig. 5.4. Now, the RGB-D videos (MVK format) are the starting point for the software. These were previously acquired in orchard environments (in static or from mobile platforms) through the AK_ACQS module that is part of the AKFruitData software [1]. Considering a frame of a set of video fragments, the software module requires certain image filtering settings to be set ('Depth and coordinates filtering' in Fig. 5.3). In this way, the user can filter a certain spatial location within the frame (filtering by coordinates) or a range of depth, discarding those objects that exceed a certain distance from the Azure Kinect camera. By making appropriate settings from the GUI, objects in the depth image can therefore be discarded by distance to obtain a 'thresholded depth image' (Fig. 5.4) which, merged with the original colour image, provides the filtered RGB image that serves as input to the detector in further processing. ‘Coordinate filtering’ will then be used to delimit the detection zone.

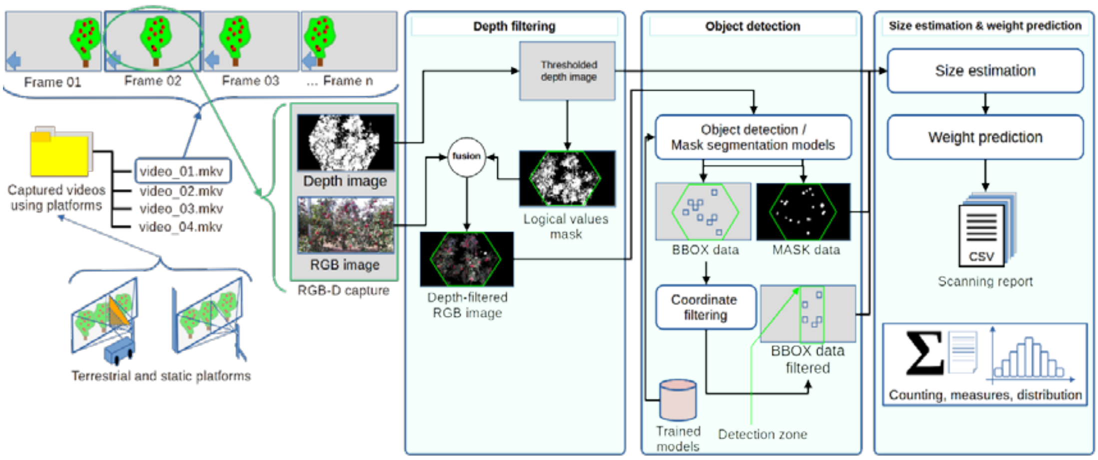


Fig. 5.4. Tasks that are performed sequentially by the AK_VIDEO_ANALYSER module that is part of the AKFruitYield dual software.

‘Object detection’ uses a trained model (Mask R-CNN or Faster R-CNN) that is applied to the depth-filtered RGB image (Fig. 5.4). As objects (fruits) are being detected, size determination (fruit axes) using the ROI data (BBOX or MASK, depending on the method) serves to then enter the geometric measurements of the apples in the allometric model. Reports in CSV format are activated at the end of the analysis. Each fruit (apple) is matched to the measurements of the major axis and the minor axis in pixels, the estimated values in mm, and the predicted weight in g.

The functionalities developed in the AK_VIDEO_ANALYSER module (Fig. 5.3) are listed below.

- “Analyse video” allows the user to configure video analysis parameters. Examples are the number of frames to analyse, filters to apply, or detection, sizing and weight prediction models to be implemented. Results are displayed on screen and conveniently organized in a CSV file.
- “Preview video” helps to configure detection zone dimensions and distance filters on colour images.

- <u>"Export frames"</u> provides the user with a set of analysed images and the information obtained. It is a useful functionality to observe how algorithms are applied on the frames.
- <u>"Reset settings"</u> allows the user default values in the GUI to be reset.
- <u>"Run in command line"</u> allows video analysis using the command line without the need for a GUI screen. Useful functionality in carrying out scriptable processes.

5.3 Illustrative examples

An example of sizing algorithm benchmarking for final yield prediction is shown in Fig. 5.5. First (Fig. 5.5a), selection of the appropriate options in the GUI of the AK_SW_BENCHMARKER module causes binary masking to achieve ROIs of labelled apples to be bounded, to then fit ellipses (EF) to the ROIs over the colour image. A detail of the operation for the apple with label 2136 is shown in Fig. 5.5b. Red points on the edge of the ROI serve as a guide for fitting the ellipse (in green) and obtaining the semi-major and semi-minor axes of the apple marked in blue and green, respectively.

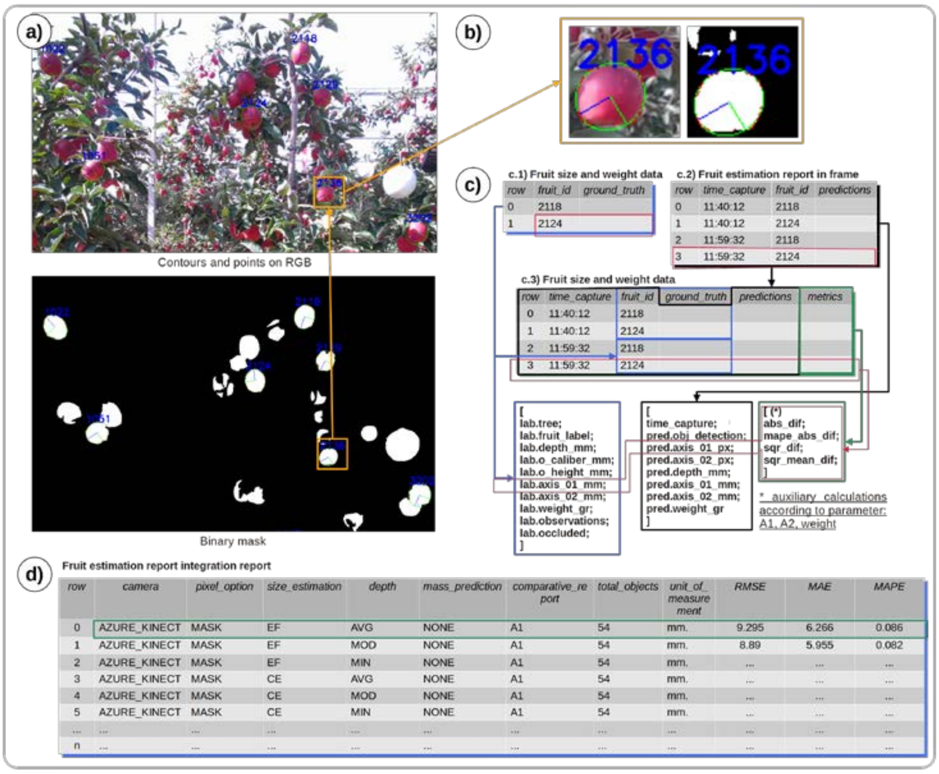
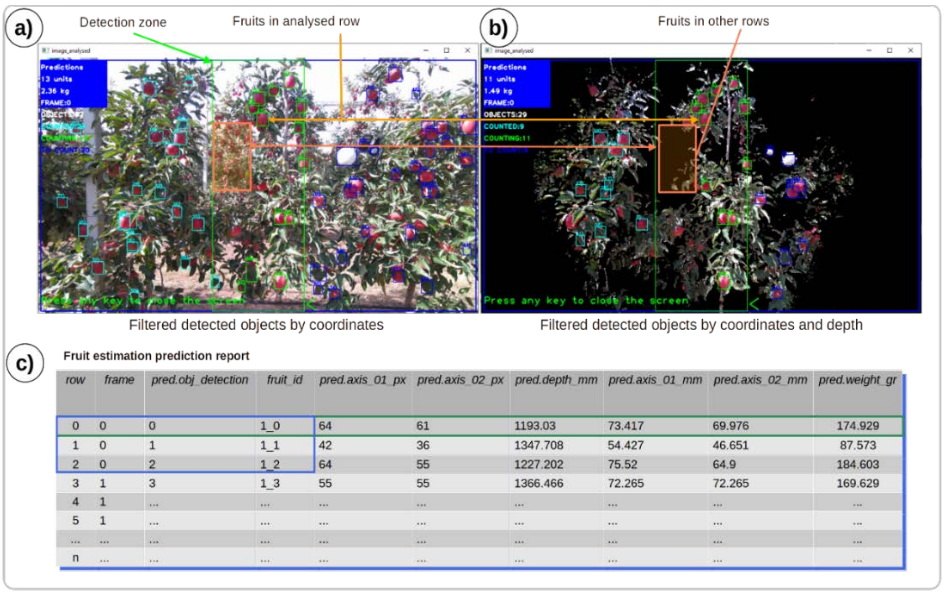


Fig. 5.5. Example of the operation of the functionalities of the AK_SW_BENCHMARKER module.

Extracts in Fig. 5.5c work as follows. Ground truth data of size and weight per labelled apple (c.1) are shown together with the intermediate results (prediction data, c.2). Final report (c.3) includes the compared ground truth and predicted data of geometric length (major axis 01 or minor 02) or weight at the request of the software user. The last piece of information in Fig. 5.5d shows an overview report. In the case of the figure, the user can view the results of the error metrics which, by assessing the major axis (A1), have been obtained according to different sizing and depth algorithms on the set of labelled apples for which real size data are available. In this way, the user has comparative information between methods (algorithms). It is worth mentioning that preliminary tests with the AK_SW_BENCHMARKER have provided fruit size estimates (non-occluded fruits) with MAPE < 5%. These results values are comparable to those obtained with other state-of-the-art 3D sensing techniques [9]. The AK_SW_BENCHMARKER have allowed to identify the best combinations of sizing algorithms and allometric models for which fruit weight predictions with a MAPE < 6 % were achieved, lower than the 10 % of error threshold usually accepted in yield predictions.

As for the other software, the AK_VIDEO_ANALYSER module, some of its features can be seen in Fig. 5.6. Use of the module is initially argued for analysis of videos recorded with the Azure Kinect camera on trees within an apple orchard. However, the module has the possibility of being adapted and used in other orchards if detection algorithms and weight prediction models are made available.



row	frame	pred.obj_detection	fruit_id	pred.axis_01_px	pred.axis_02_px	pred.depth_mm	pred.axis_01_mm	pred.axis_02_mm	pred.weight_gr
0	0	0	1_0	64	61	1193.03	73.417	69.976	174.929
1	0	1	1_1	42	36	1347.708	54.427	46.651	87.573
2	0	2	1_2	64	55	1227.202	75.52	64.9	184.603
3	1	3	1_3	55	55	1366.466	72.265	72.265	169.629
4	1	...	...	...	...	...	...	...	...
5	1	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
n	...	...	...	...	...	...	...	...	...

Fig. 5.6. Example of some functionalities of the AK_VIDEO_ANALYSER module.

Capture of three apple trees in a given row are shown (Fig. 5.6a). Note that apples belonging to trees in the back row are detected (orange box) by coordinates and distance filtering (Fig. 5.6b) and are not counted. This is a very useful functionality to delimit the fruits to be detected and processed, avoiding errors in final yield predictions at a spatial plot scale. As the AK-VIDEO_ANALYSER module works, three detection zones are shown to the user corresponding to: i) fruits already counted (in light blue); ii) fruits being counted and real use phase of sizing and weight algorithms (in light green); and iii) detected fruits that will be analysed shortly (in dark blue).

Real-time informative data is displayed in the upper left-hand part of the screen. At the end of the analysis, a report (information shown in Fig. 5.6c) summarizes size (in pixels and mm), distance location (depth in mm) and weight (in g) for each detected fruit. In addition to these tests, algorithms have been applied to foam spheres (visible in Fig. 5.6). Consequently, different examples of apples and foam spheres are included in the software repository to check the applicability of the AK_VIDEO_ANALYSER software module on two different types of objects.

5.4 Impact

With the development of the AKFruitYield software (AK_SW_BENCHMARKER and AK_VIDEO_ANALYSER modules), functionalities of other related and open access software (AKFruitData [1]) are complemented. Two tools are therefore made available to the user to allow the use of RGB-D Azure Kinect cameras to, in a first step, acquire data in orchards and create analysable datasets (AKFruitData) and, in a second step, to process the data (videos and images) and provide reliable fruit yield predictions (AKFruitYield).

The impact of AKFruitYield (in combination with AKFruitData) is expected to be important for two main reasons: i) by providing fruit growers and, especially, managers in fruit growing an application that can significantly reduce the time and cost of a task still carried out manually in many farms; and ii) by promoting the introduction of low-cost accessible technologies under a strategic framework of digitizing the fruit sector.

AKFruitYield is open source software designed to allow future updates. New detectors can be implemented, sizing algorithms improved and allometric models refined as its use adds more knowledge and users demand new specifications (other apple tree cultivars, including other fruit species, or applications in earlier stages of fruit development as a scouting tool). For the present, AKFruitYield provides a practical solution in terms of yield prediction (and not simple fruit counting) in apple orchards. Additionally, a powerful benchmarking module is made available to test other detection, sizing and yield allometry options that may be implemented as a result of its use. Multi-Object Tracking and Segmentation (MOTS) [29] is another powerful option for greater precision that could be implemented in the future, once testing and the required adaptations were made in the current version of the software.

AKFruitYield (presented here) and AKFruitData [1] have been designed to encourage the use of the Azure Kinect camera (and depth cameras in general) by end users working or doing research in fruit growing. Ease of use is a key aspect and, for this reason, video examples and image datasets to verify the operation of the software are attached as supplementary material to the source code. It is possible to use this software with RGB-D sensors other than the Azure Kinect camera. However, implementing specific data extraction routines (wrappers) according to the selected devices would be required.

5.5 Conclusions

AKFruitYield is presented in a modular format since it includes two separate but complementary modules focused on processing fruit data acquired through RGB-D Azure Kinect cameras in apple orchards. The first module (AK_SW_BENCHMARKER) makes it possible to apply different sizing algorithms and allometric yield prediction models on colour and depth tree images containing previously detected and manually labelled apples. The second module (AK_VIDEO_ANALYSER) is a video analysis software that allows automatic apple detection, size estimation and yield prediction to be performed by applying the best ranked algorithms resulting from the previous benchmarking. A special objective has been to develop easy-to-use graphical interfaces for end users working in research as well as fruit growers and advisory technicians who require new digital tools for better management of fruit farms.

Future works are planned to expand the current functionalities and provide support under the concept of continuous development and improvement. In short, with the AKFruitYield software, the cycle that started with the AKFruitData software is successfully completed, thus meeting the initial objectives of acquiring and processing data from Azure Kinect cameras in apple orchards for yield prediction.

5.6 Declaration of competing interest

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

5.7 Acknowledgements

This work was partly funded by the Department of Research and Universities of the Generalitat de Catalunya (grants 2017 SGR 646) and by the Spanish Ministry of Science and Innovation/AEI/10.13039/501100011033/ERDF (grant RTI2018-094222-B-I00 [PAgFRUIT project] and PID2021-126648OB-I00 [PAgPROTECT project]). The Secretariat of Universities and Research of the Department of Business and Knowledge of the Generalitat de Catalunya and European Social Fund (ESF) are also thanked 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. The authors would also like to thank the Institut de Recerca i Tecnologia Agroalimentàries (IRTA) for allowing the use of their experimental fields, and in particular Dr. Luís Asín and Dr. Jaume Lordán who have contributed to the success of this work.

5.8 References

[1] Miranda JC, Gené-Mola J, Arnó J, Gregorio E. AKFruitData: A dual software application for Azure Kinect cameras to acquire and extract informative data in yield tests performed in fruit orchard environments. SoftwareX. 2022; 20:101231. https://doi.org/10.1016/j.softx.2022.101231

[2] Bargoti S, Underwood JP. Image segmentation for fruit detection and yield estimation in apple orchards. J F Robot. 2017; 34(6):1039-1060. https://doi.org/10.1002/rob.21699

[3] He L, Fang W, Zhao G, Wu Z, Fu L, Li R, Majeed Y, Dhupia J. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. Comput Electron Agric. 2022; 195(February):106812. https://doi.org/10.1016/j.compag.2022.106812

[4] Zhang C, Valente J, Kooistra L, Guo L, Wang W. Orchard management with small unmanned aerial vehicles: a survey of sensing and analysis approaches. Precis Agric. 2021; 22(6):2007-2052. https://doi.org/10.1007/s11119-021-09813-y

[5] Conti G, Mercante E, de Souza EG, Sobjak R, Bazzi CL. AGDATABOX-RS computational application: Remote sensing data management. SoftwareX. 2023; 23:101435. https://doi.org/10.1016/j.softx.2023.101435

[6] Correndo AA, Pearce A, Bolster CH, Spargo JT, Osmond D, Ciampitti IA. The soiltestcorr R package: An accessible framework for reproducible correlation analysis of crop yield and soil test data. SoftwareX. 2023; 21:101275. https://doi.org/10.1016/j.softx.2022.101275

[7] Bazzi CL, Jasse EP, Graziano Magalhães PS, Michelon GK, de Souza EG, Schenatto K, Sobjak R. AgDataBox API – Integration of data and software in precision agriculture. SoftwareX. 2019; 10:100327. https://doi.org/10.1016/j.softx.2019.100327

[8] MacEachern CB, Esau TJ, Schumann AW, Hennessy PJ, Zaman QU. Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks. Smart Agric. Technol. 2023; 3:100099. https://doi.org/10.1016/j.atech.2022.100099

[9] Miranda JC, Gené-Mola J, Zude-Sasse M, Tsoulias N, Escolà A, Arnó J, Rosell-Polo JR, Sanz-Cortiella R, Martínez-Casasnovas JA, Gregorio E. Fruit sizing using AI: A review of methods and challenges. Postharvest Biol. Technol. 2023; 206:112587. https://doi.org/10.1016/j.postharvbio.2023.112587

[10] Parr B, Legg M, Alam F. Grape yield estimation with a smartphone's colour and depth cameras using machine learning and computer vision techniques. Comput. Electron. Agric. 2023; 213:108174. https://doi.org/10.1016/j.compag.2023.108174

[11] Gregorio, E., Llorens, J. (2021). Sensing Crop Geometry and Structure. In: Kerry, R., Escolà, A. (eds) Sensing Approaches for Precision Agriculture. Progress in Precision Agriculture. Springer, Cham. https://doi.org/10.1007/978-3-030-78431-7_3

[12] Condotta ICFS, Brown-Brandl TM, Pitla SK, Stinn JP, Silva-Miranda KO. Evaluation of low-cost depth cameras for agricultural applications. Comput Electron Agric. 2020; 173:105394. https://doi.org/10.1016/j.compag.2020.105394

[13] Pasinetti S, Nuzzi C, Luchetti A, Zanetti M, Lancini M, De Cecco M. Experimental procedure for the metrological characterization of time-of-flight cameras for human body 3D measurements. Sensors. 2023; 23(1):538. https://doi.org/10.3390/s23010538

[14] Gené-Mola J, Llorens J, Rosell-Polo JR, et al. Assessing the Performance of RGB-D sensors for 3D Fruit crop canopy characterization under different operating and lighting conditions. Sensors. 2020; 20(24):7072. https://doi.org/10.3390/s20247072

[15] Lopes A, Souza R, Pedrini H. A survey on RGB-D datasets. Comput. Vis. Image Underst. 2022; 222:103489. https://doi.org/10.1016/j.cviu.2022.103489

[16] Hacking C, Poona N, Manzan N, Poblete-Echeverría C. Investigating 2-D and 3-D proximal remote sensing techniques for vineyard yield estimation. Sensors. 2019; 19(17):3652. https://doi.org/10.3390/s19173652

[17] Wang W, Li C. Size estimation of sweet onions using consumer-grade RGB-depth sensor. J Food Eng. 2014; 142:153-162. https://doi.org/10.1016/j.jfoodeng.2014.06.019

[18] Wang Z, Walsh KB, Verma B. On-tree mango fruit size estimation using RGB-D images. Sensors (Switzerland). 2017; 17(12):1-15. https://doi.org/10.3390/s17122738

[19] Sun M, Xu L, Luo R, Lu Y, Jia W. Fast location and recognition of green apple based on RGB-D image. Front Plant Sci. 2022; 13. https://doi.org/10.3389/fpls.2022.864458

[20] Liu J, Yuan Y, Zhou Y, Zhu X, Syed T. Experiments and analysis of close-shot identification of on-branch citrus fruit with RealSense. Sensors. 2018; 18(5):1510. https://doi.org/10.3390/s18051510

[21] Xiong Y, Peng C, Grimstad L, From PJ, Isler V. Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. Comput Electron Agric. 2019; 157(January):392-402. https://doi.org/10.1016/j.compag.2019.01.009

[22] Gené-Mola J, Vilaplana V, Rosell-Polo JR, Morros J-R, Ruiz-Hidalgo J, Gregorio E. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. Comput Electron Agric. 2019; 162:689-698. https://doi.org/10.1016/j.compag.2019.05.016

[23] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE; 2017:2980-2988. https://doi.org/10.1109/ICCV.2017.322

[24] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell. 2017; 39(6):1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[25] Miranda JC, Gené-Mola J, Arnó J, Gregorio E. AK_SM_RECORDER, https://pypi.org/project/ak-sm-recorder/, 2022 [accessed 26 September 2023].

[26] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis. 2010; 88(2):303-338. https://doi.org/10.1007/s11263-009-0275-4

[27] The Matroska Association. Matroska, https://www.matroska.org/, [accessed 26 September 2023].

[28] Lin T-Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, eds. Computer Vision -- ECCV 2014. Springer International Publishing; 2014:740-755. https://doi.org/10.1007/978-3-319-10602-1_48

[29] de Jong S, Baja H, Tamminga K, Valente J. APPLE MOTS: Detection, segmentation and tracking of homogeneous objects using MOTS. IEEE Robot. Autom. Lett. 2022; 7:11418–11425. https://doi.org/10.1109/LRA.2022.3199026

Chapter 6: Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples

This chapter was published as:
Miranda, J.C., Arnó, J., Gené-Mola, J., Lordan, J., Asín, L., Gregorio, E., 2023. Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples. Comput. Electron. Agric. 214, 108302. https://doi.org/https://doi.org/10.1016/j.compag.2023.108302

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

Juan C. Miranda[1]*, Jaume Arnó[1], Jordi Gené-Mola[1,2], Jaume Lordan[3], Luis Asín[3], Eduard Gregorio[1]

[1]Research Group in AgroICT & Precision Agriculture, Department of Agricultural and Forest Sciences and Engineering, Universitat de Lleida - Agrotecnio CERCA Centre, 25198, Lleida, Catalonia, Spain

[2]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.

[3]Fruit Production Program, Institute of AgriFood, Research and Technology (IRTA), Parc Científic i Tecnològic Agroalimentari de Gardeny (PCiTAL), Fruitcentre, 25003 Lleida, Spain.

*Corresponding author: juancarlos.miranda@udl.cat (J.C. Miranda)

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 (R2 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.

Keywords: Azure Kinect, fruit sizing, allometric weight models, apple tree, digital fruit growing.

6.1 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 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.

6.2 Materials and methods

Fig. 6.1 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. 6.1). 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. 6.1). 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. 6.1). 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. 6.1). 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. 6.1). In a final step, the performance of the proposed algorithms was evaluated by contrasting several different statistical metrics ('evaluation and testing', Fig. 6.1).

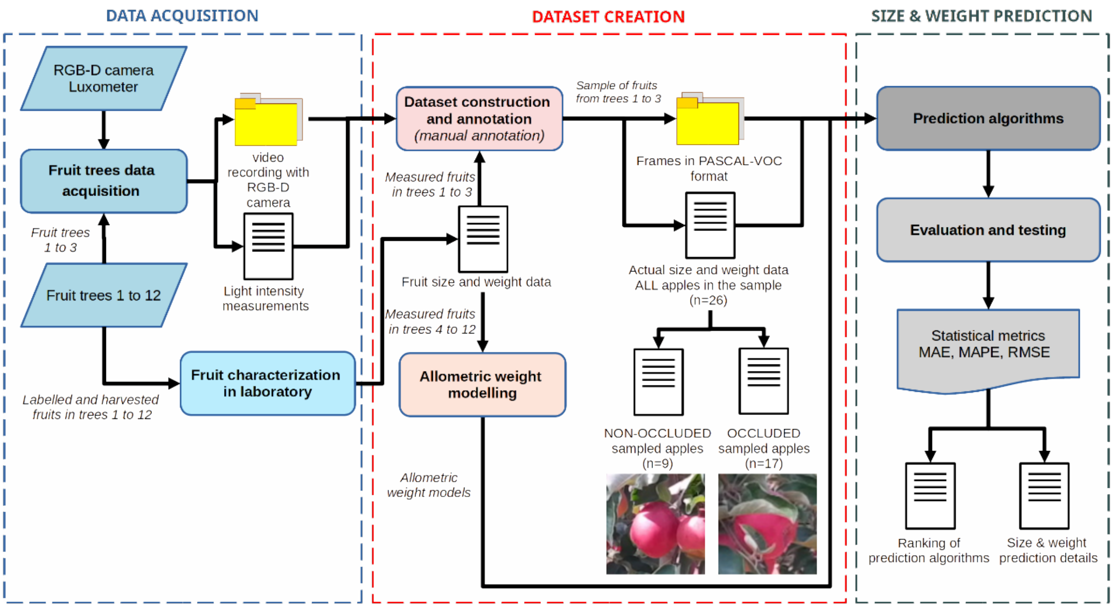


Fig. 6.1. Sequential methodology for the prediction of apple fruit size and weight using data collected with an RGB-D camera under field conditions. The three blocks (delimited by dotted lines) make up the global procedure, with

each one including the different steps performed (rounded boxes) and highlighting the input and output data required and provided in each case.

6.2.1 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. 6.2a,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.

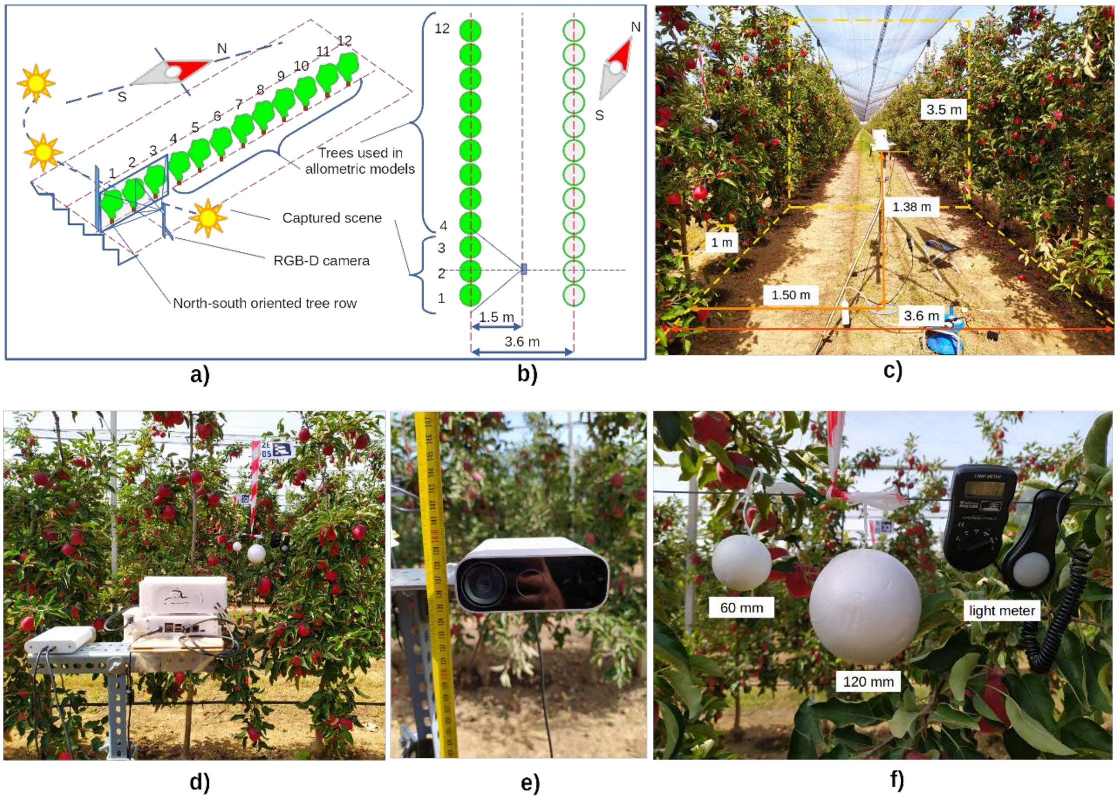


Fig. 6.2. Field experimental set-up. a) A perspective representation of the scene, showing the relative position of the sun throughout the experiment. b) Plan view of the layout. c) View of the tree alley, showing the planting pattern and camera position. d) View of the captured scene (trees 1 to 3). e) Azure Kinect camera. f) Calibration foam spheres and digital light meter placed on the trees.

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 2.1 (Microsoft, 2022).

Table 6.1. 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

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. 6.2c,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. 6.2f, 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 8.2º Brix, and firmness values of 14.6 kg/cm2. A total of 25 video recordings (one every 15-20 minutes), each with a duration of 4 seconds, 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. 6.3a), using adhesive paper stickers (Fig. 6.3b). Several videos, identification notes and photos were also taken as ancillary data, in order to keep evidence of fruit positions within each tree (Fig. 6.3c). The fruits were hand-picked on September 30 and October 1, 2021 and placed in different collapsible plastic storage boxes (one for each tree).

a) b) c)

Fig. 6.3. 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.

6.2.2 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. 6.4a).

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. 6.4b). A CB 501 weighing machine (Adam Equipment, Oxford, CT, USA) was used to measure fruit weight (W) (Fig. 6.4c). 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.

a) b) c)

Fig. 6.4. 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.

6.2.3 *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 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: $W = \beta_0 + \beta_1 (D_1^2 D_2) + \varepsilon$, and $W =$

$\beta_0 + \beta_1(D_1 D_2^2) + \varepsilon$. 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 R2 in each case. To avoid relying solely on R2 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. 6.1 and 6.2). 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.

6.2.4 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. 6.5). 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.

Fig. 6.5. 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.

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 8T 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 6.2.2) and the fruits identified in the labelling process (Section 6.2.4), 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. 6.1, apples within the dataset were grouped into two subsets, non-occluded apples (n=9) and occluded apples (n=17).

6.2.5 Prediction algorithms

Fig. 6.6 provides an overview of the algorithms for fruit size and weight prediction developed in this work. Image datasets in PASCAL-VOC format (Section 6.2.4), which include RGB images, depth images and binary masks, were used as input for the prediction algorithms (Fig. 6.6a). Two different approaches were then used to identify the regions of interest (ROI): i) bounding boxes (BBOX) (Fig. 6.6b.1); and ii) binary masks (MASK) (Fig. 6.6b.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 6.2.3 were applied to predict fruit weight (Fig. 6.6c).

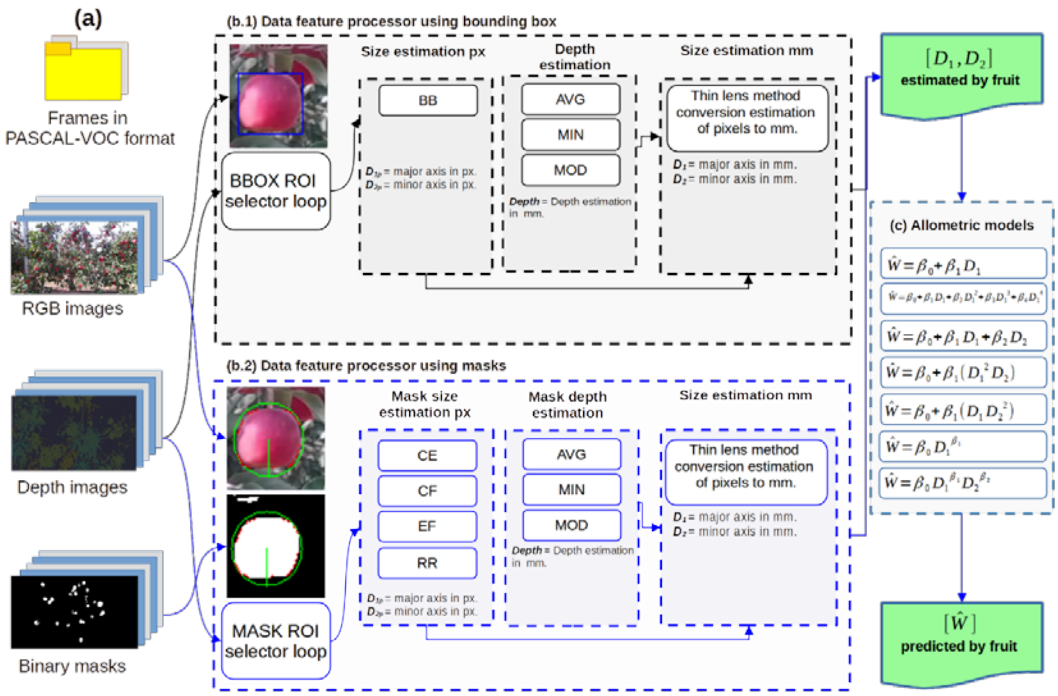


Fig. 6.6. Overview of the size and weight prediction algorithms applied. a) Input for prediction algorithms. Approaches for identifying regions of interest (ROI): b.1) bounding box (BBOX), b.2) binary masks (MASK). Size estimation: bounding box (BB), circle enclosing (CE), circle fitting (CF), ellipse fitting (EF), rotated rectangle (RR). Depth estimation: average (AVG), minimum (MIN), modal (MOD).

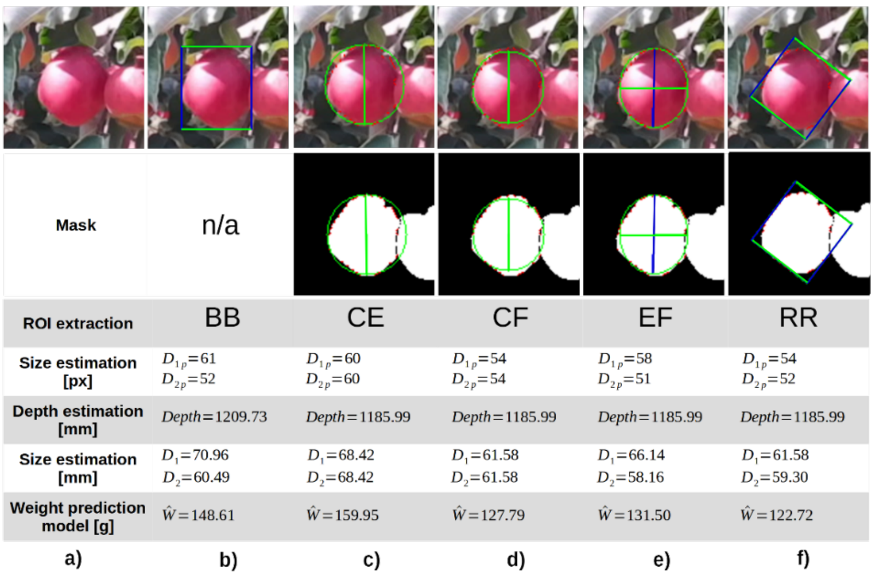
The prediction algorithms were implemented using Python 3.8, Tkinter, and OpenCV for image processing, and also other open source libraries/packages, such as Numpy, Scikitlearn and Pandas. As result, a software package with graphic user interfaces was published as Python package containing the pipeline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

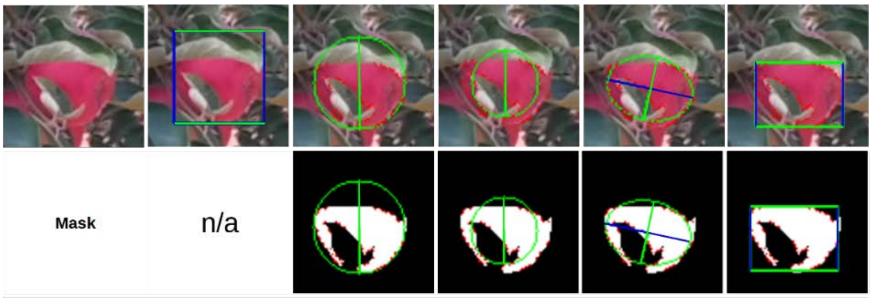
6.2.5.1 Size estimation in pixels

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. 6.7). 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. 6.7).

- 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.





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. 6.7. Size and weight estimates of: (TOP) non-occluded apple # 2167, $C = 63.68\ mm$, $H = 66.07\ mm$, $W = 136.4\ g$; (BOTTOM) occluded apple # 2171, $C = 64.43\ mm$, $H = 54.57\ mm$, $W = 102.3\ g$. 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 6.3.

6.2.5.2 Depth estimation

To estimate actual fruit sizes (in mm) from measurements in pixels (Section 6.2.5.1), 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. 6.7 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. 6.8a represents the regions used for depth estimation (RGB image and binary mask), while Fig. 6.8b 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.

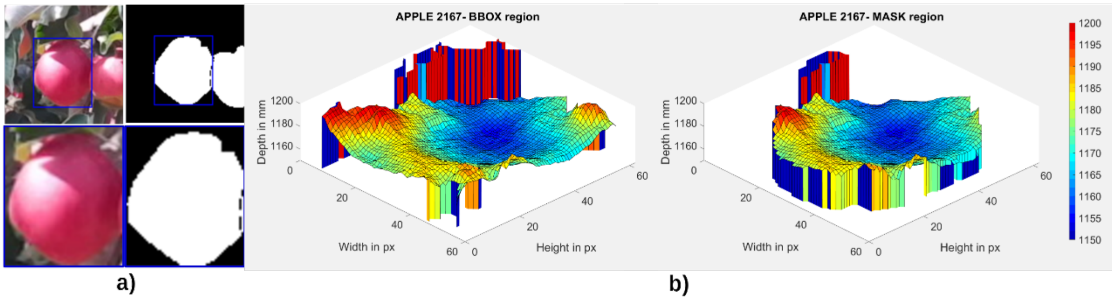


Fig. 6.8. Depth estimation of apple #2167 (taken at 12:35 UTC+2), $BBOX\ average\ depth = 1209.73\ mm$, $MASK\ average\ depth = 1185.99\ mm$. a) RGB image and binary mask selected by bounding box rectangle. b) Depth values within the bounding box and mask region.

6.2.5.3 Estimations of fruit size and predictions of fruit weight

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

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

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 6.3.2) to predict fruit weight ($\widehat{W}$) in grams per fruit. For example, Fig. 6.7 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.

6.2.6 Evaluation and testing

The reliability of the prediction algorithms was verified using different statistical metrics. In a first step (Section 6.3.3), 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 6.3.4), 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 6.2, where $\hat{y}_i$ 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 6.2.2).

Table 6.2. Prediction algorithm evaluation metrics.

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

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. 6.6). 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.

6.3 Results

6.3.1 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. 6.9). 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 (Section 6.2.4). 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.

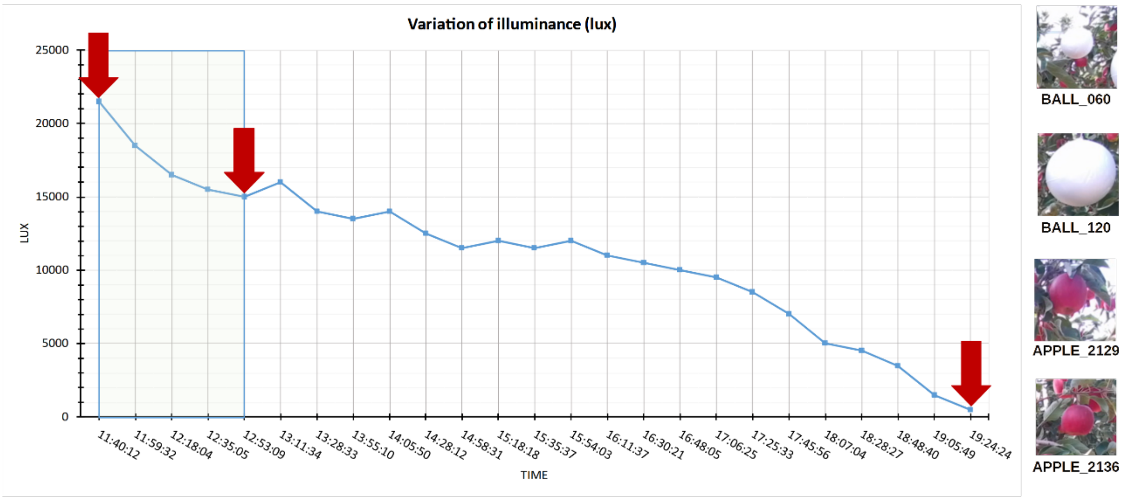
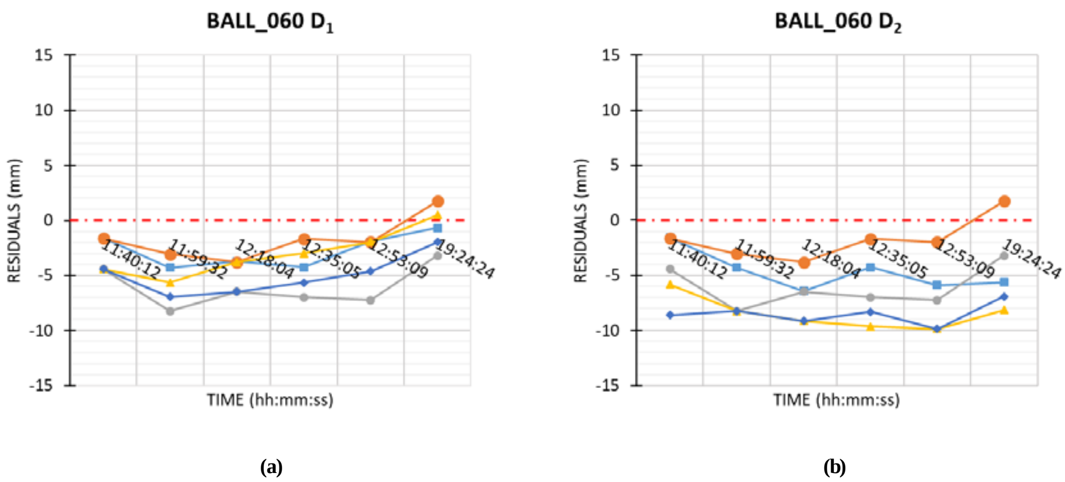
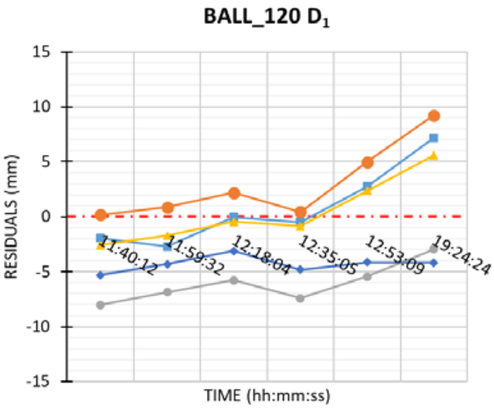


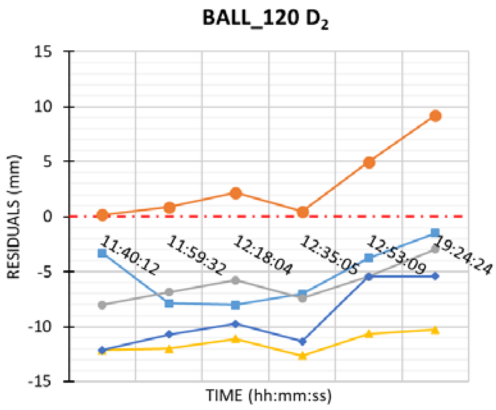
Fig. 6.9. Variation of illuminance (lux) at different times during the field data capture (September 27, 2021). The images of the spheres and apples correspond to the time slot 11:40:12 (UTC +2). Red arrows represent the moments (delimiting the capture range or a specific moment) at which measurements 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.

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. Fig. 6.10 and 6.11 (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.





(c)



(d)

Fig. 6.10. 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.*

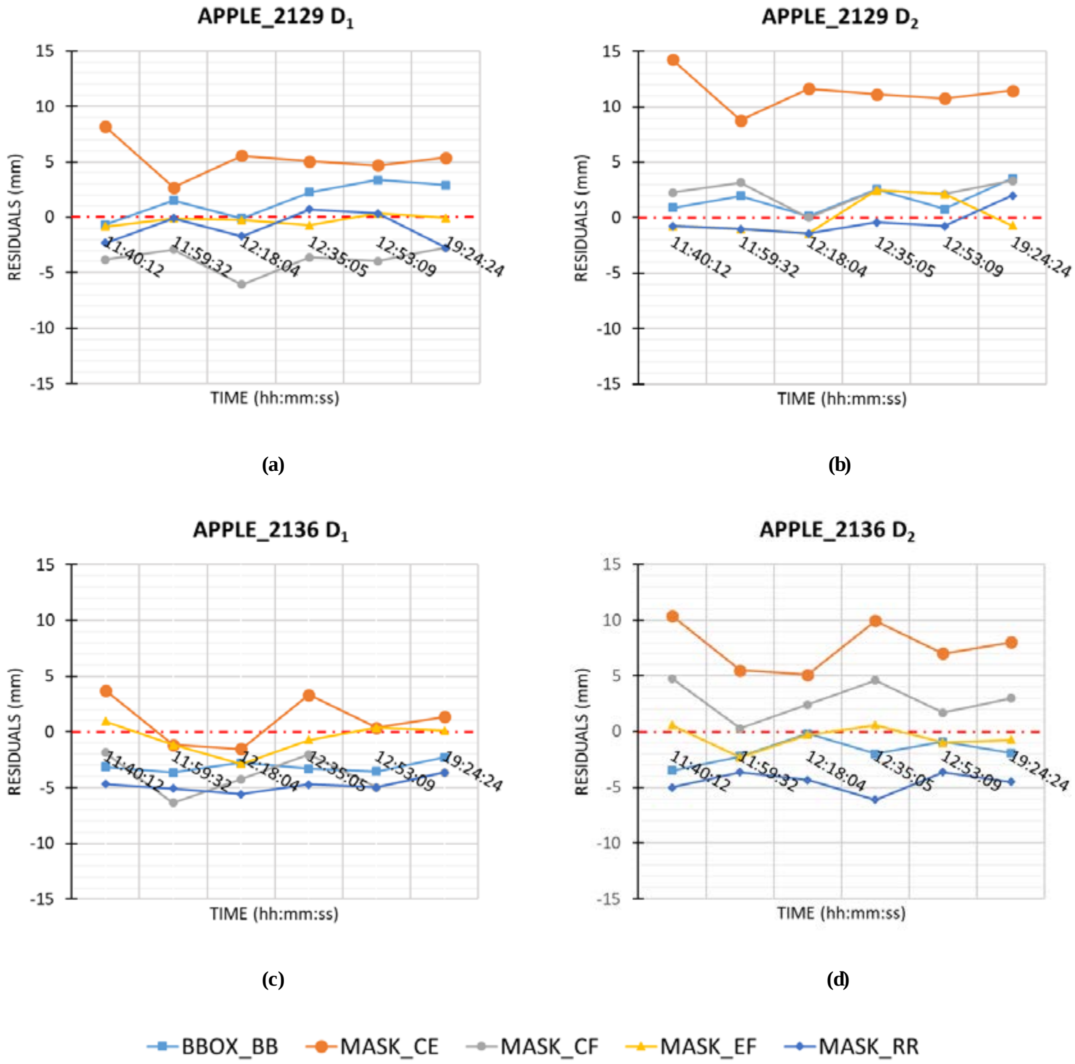


Fig. 6.11. Range of errors (estimated diameter – laboratory diameter) relating to the non-occluded apples. a-b) Apple #2129, laboratory diameter: $D_1 = 69.79$ *mm,* $D_2 = 75.87$ *mm. c-d) Apple #2136, laboratory diameter:* $D_1 = 66.27$ *mm,* $D_2 = 59.62$ *mm. The depth was estimated using the average (AVG) method.*

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. 6.10). 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. 6.11 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.

6.3.2 Allometric models for predicting apple weight

Table 6.3 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.

Table 6.3. 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

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.

6.3.3 *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 6.4. 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.

Table 6.4. 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.

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 6.4). 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. 6.12. 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 6.4 and Fig. 6.12).

In the case of the set of occluded apples (Table 6.5 and Fig. 6.13), 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. 6.13). 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 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. 6.13, 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.

Table 6.5. 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.

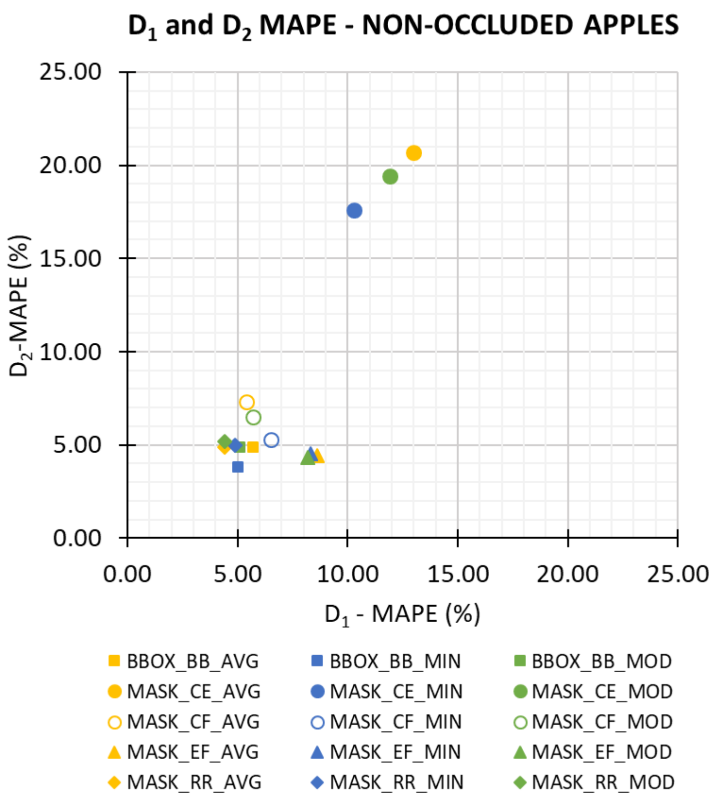


Fig. 6.12. Comparison of the Mean Absolute Percentage Error (MAPE) between D1 and D2 applied to the set of non-occluded apples (n=9).

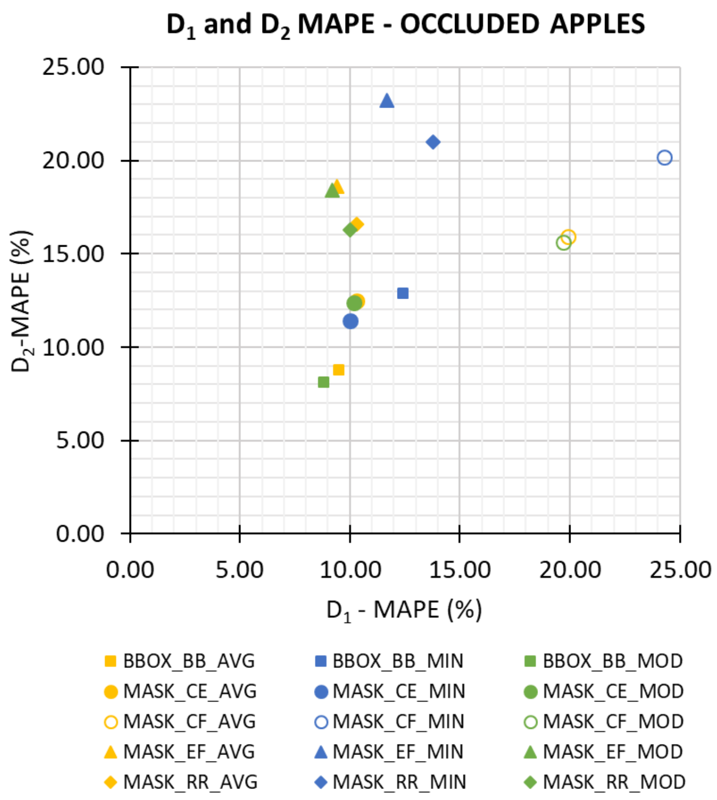


Fig. 6.13. Comparison of the Mean Absolute Percentage Error (MAPE) between D1 and D2 applied to the set of occluded apples (n=17).

6.3.4 Optimal combined sizing algorithm and allometric model for predicting apple weight

For non-occluded fruits (Table 6.6), sizing the apples using circle fitting (CF) and the subsequent application of the linear allometric model (3) (Table 6.3, Section 6.3.2) 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 6.6), 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 6.6).

Table 6.6. 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 6.3.

Somewhat surprisingly, the sizing methods that provided best results in terms of estimating apple size (Section 6.3.3), 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 6.3.3), 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 6.3. The resulting prediction errors are shown in Table 6.7.

Table 6.7. 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 6.3.

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 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.

6.4 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 lux 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 lux (Fig. 6.9), using an Azure Kinect camera. No significant differences were appreciated in size estimates compared to those registered in the late afternoon (500 lux) (Fig. 6.10 and Fig. 6.11). These results indicate that the Azure Kinect was not significantly influenced by 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 6.3.3. The estimation errors for non-occluded apples (Table 6.4: 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 6.5: 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 6.3.4), accurate estimates were obtained for non-occluded apples (Table 6.6, 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 6.6) 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 of lower computational cost and more direct integration with current object detectors.

In the case of allometric models (Table 6.7), 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 6.6). 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 6.4; Fig. 6.12) are also valid options and can complement the allometric model and predict yield in an acceptable way (Table 6.6). 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.

6.5 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.

6.6 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) and 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 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.

6.7 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., 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

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., 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., 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

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., 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

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., 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

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

Chapter 7: Discussion

In this chapter, a general discussion of the doctoral thesis is conducted. The detailed discussion of each of the works that conform this thesis can be found in Chapters 3, 4, 5 and 6.

7.1 Progress in fruit detection and sizing using AI

In the first part of the thesis (Chapter 3), a review of methods for detecting/counting and measuring fruits on the tree has been carried out. In relation to fruit counting, sensors for automatic detection have been used for more than 30 years (Plá et al., 1993). With an initial research focus on the development of algorithms for object detection in images (Gongal et al., 2015), artificial intelligence has recently allowed digital image processing systems with pattern recognition to evolve towards systems based on deep learning. Progress in this respect has been impressive, with detection systems currently reaching F1-scores higher than 90% (Table 3.1; Chapter 3) and similar to the human eye. However, some important issues remain pending such as fruit tracking (required to not double-count fruits detected from different sensor positions) and estimation of quality parameters (for better harvesting management). Therefore, new research transfer opportunities are emerging with the priority of developing robust and low-cost systems, and also under the premise of having to process high amounts of data when applied to farms with a large number of trees and a large load of fruits.

Regarding the in-field fruit size estimation, two methodological approaches have been identified (Section 3.4): those based on 2D images and those that use 3D point clouds. 2D image acquisition (RGB, thermal) can be carried out using low-cost cameras. However, calibration of these images is required (e.g. using targets of known dimensions located in the same plane of the fruit to be measured), which restricts their practical application to a few trees. On the other hand, the use of 3D sensing systems such as photogrammetric techniques, LiDAR or RGB-D sensors, allows three-dimensional reconstruction (point cloud) of the orchard and direct measurement of the fruits. Fruit sizing is at a less advanced development stage than fruit counting, although estimates with a mean absolute error (MAE) of less than 4 mm (Table 3.4; Chapter 3) have been achieved. In this thesis, the use of RGB-D sensors has been chosen (Chapters 4, 5 and 6), which simultaneously provide colour and depth data. These sensors have a lower cost than LiDAR technology, and reduced computational load compared to photogrammetric techniques such as Multi-View Stereo or Structure-from-Motion. Specifically, a time-of-flight RGB-D sensor (i.e. Microsoft Azure Kinect DK) has been selected due to its greater robustness under natural lighting compared to RBG-D sensors based on structured light or active stereovision (Gregorio and Llorens, 2021).

From the experience and works analysed in Chapter 3, it would be extremely useful for the scientific community to share the datasets, algorithms and models used in fruit detection. The use of standard metrics to make methods and results objectively comparable is also necessary and, in our opinion, the F1-score for object detection and the MAE (or MAPE) for size estimation are recommended. At the same time, there is a need to continue research in computer vision and deep learning techniques applied to fruit growing early in the season, especially in the initial phases of fruit growth (before or after thinning) and even in flower count (Aggelopoulou et al., 2010).

The review work presented in Chapter 3 serves as introduction for researchers who wish to delve into deep learning techniques and their application to fruit detection and sizing. Foundations of convolutional neural networks (CNNs) and main types (image classification CNNs, object detection CNNS, semantic and instance segmentation CNNs) are presented. Moreover, its didactic approach, together with the special emphasis on fruit size and maturity estimation methodologies, represent the main contributions of this chapter to the current literature on fruit detection.

7.2 Software development for RGB-D cameras in fruit orchards

The second part of this thesis (Chapters 4 and 5) is focused on the development of software tools to use the Azure Kinect RGB-D camera in fruit orchard environments. In Chapter 4, a dual software called AkFruitData has been presented, which allows the simultaneous acquisition of data with different sensors (AK_ACQS) and the extraction of datasets (AK_FRAEX) from the videos acquired by the Azure Kinect camera. In Chapter 5, the AKFruitYield software has been introduced, which includes two modules, the AK_SW_BENCHMARKER and the AK_VIDEO_ANALYSER. The first of these modules allows the application of different fruit sizing and fruit weighting prediction algorithms, while the second is used to count fruits (apples) and predict their yield at plot scale from videos captured by the Azure Kinect.

The software tools developed in this thesis cover all stages ranging from the in-field data acquisition to the fruit yield prediction. Although the software has been developed for primary use with the Azure

Kinect camera, it can be adapted to be applied to other RGB-D sensor models. Furthermore, as it is a modular software, it can also be used when only one of the specific functionalities is required: data acquisition, data extraction, fruit detection, size and weight estimation, or yield prediction.

From the point of view of the software developer, the selection of the Python's ecosystem and the use of open source libraries has the advantage of frequent updates by the community. On the other hand, changing versions can lead to incompatibilities between different components of the developed software, which translates into a certain fragility of the system.

To the author's knowledge, this is the first software published in a peer-reviewed, open-access journal where fruit yield prediction is addressed using a 3D sensing system. One of the aims of this work is to encourage other precision agriculture software developers to publish their work as journal articles, which would allow a more rigorous software benchmarking. It should be noted that all the software code developed in the framework of this thesis can be found in open-access repositories, hoping its adoption can be useful to farmers, technical advisors and researchers.

7.3 Fruit sizing estimation and weight prediction

In the third part of this thesis (Chapter 6), experimental works are presented, as well as the fruit size and weight predictions resulting from them. Static measurements of apple trees were performed on an experimental farm using the Azure Kinect camera together with the AKFruitData software presented in Chapter 4. Subsequently, fruit characterization (geometric dimensions and weight) of harvested apples was made in the laboratory. These data were used to develop seven allometric models (Table 6.3; Chapter 6). Using the AK_SW_BENCHMARKER software presented in Chapter 5, different combinations of algorithms were evaluated to predict the size and apple fruit weight from the acquired colour and depth images. For non-occluded apples, size estimates with MAE values between 3-3.5 mm were obtained (Table 6.4; Chapter 6), similar to those currently present in the state of the art (Table 3.4; Chapter 3). Regarding fruit weight predictions, mean absolute percentage errors (MAPE) of less than 6% were achieved (Table 6.6).

Nowadays, uncertainties of up to 10% in commercial yield predictions are accepted. As seen, results of this thesis clearly overcome this threshold, opening the door to RGB-D sensors as an alternative to manual-based fruit sizing and yield prediction methods. Although future work should include in-field validations with a greater number of trees (Section 7.4), the potential of this methodology is shown both in terms of cost savings (lower labour and temporal costs) and in terms of highly accurate estimates (no subjective appreciation errors).

The developed methodology also presents its own limitations, particularly when measuring occluded apples. As shown in Tables 6.5 and 6.6 (Chapter 6), size estimates in occluded apples presented a MAE of between 6-8 mm, while the MAPE in weight predictions exceeded 18%. The application of amodal instance segmentation (Gené-Mola et al, 2023), which allows reconstruction of the occluded parts of the fruit, may be a way to improve these results. Alternatively, occluded apples could be discarded in the final estimates, thus minimizing the uncertainties in yield prediction due to them (Neupane et al., 2022). Typically, a limiting factor when using RGB-D cameras has been their poor performance under sunlight. The results of this thesis demonstrate that the current generation of time-of-flight RGB-D cameras (Azure Kinect) allows accurate measurements (Figure 6.10 and 6.11; Chapter 6) to be carried out in high illuminance conditions (higher than 15000 lux).

Allometric models used to predict the weight of apples (Chapter 6) deserve a more detailed explanation. In the process of obtaining these models, it is worth commenting on a typical problem that normally appears with predictors (at least when dealing with linear models). This problem refers to collinearity in the case of the linear model that uses the geometric measurements of apple axes D_1 and D_2 as separate but additive predictors. In fact, this model has turned out to be the most accurate when estimating the weight of apples (Table 6.6, Chapter 6). As mentioned in Faraway (2016), the effect of collinearity can be summarized by the formula,

$$var\hat{\beta}_i = \sigma^2 \left(\frac{1}{1-R_i^2}\right) \frac{1}{\sum_j (D_{ij}-\overline{D}_i)^2} \quad (1)$$

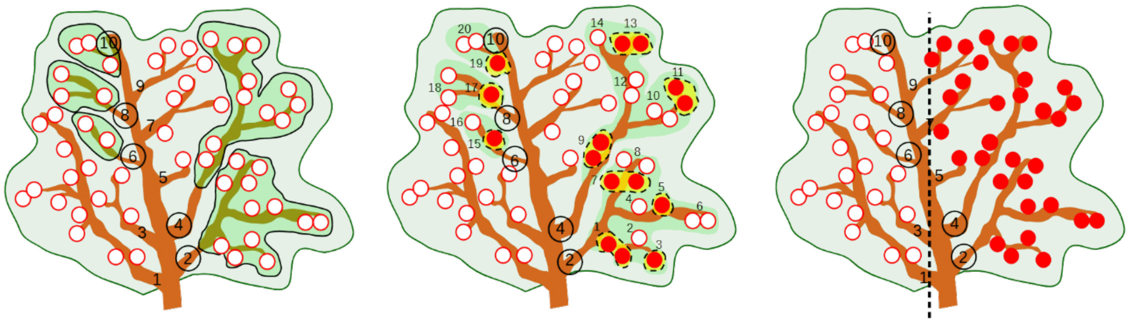
with R_i^2 the coefficient resulting from the regression of one axis on the other axis. Since a high correlation between axes (R_i^2 close to one) is expected, this collinearity would provide imprecise estimates of β. Faced with this problem, all that remained was to take a sample of apples with a very variable range of sizes, spreading the values of the axes as much as possible as has been the case in the research of this part of the thesis. Avoiding the problem of collinearity also involves the option of discarding one of the predictors. This has been the idea of adjusting a simple regression model using as a predictor the major axis D_1 of the apple (normally coinciding with the so-called size of the fruit).

The compensation between the sizing error and that committed in allometric modelling deserves special mention. Obviously, an additive compensating effect has occurred. But, this is a consequence of the proposed procedure. Error in sizing the apples is introduced into an allometric model that also contains its own error term. The result, unpredictable a priori, can lead to a very favourable final weight estimation error. Interestingly, the allometric models that provided a greater estimation error (RMSE) when they were obtained, have finally been the models with the best performance using the sizes (axes) estimated by the sizing algorithms. Probably, they have been the models that have applied a greater and more timely compensation effect.

7.4 Future works

In the short term, it is planned to apply the RGB-D based methodology developed in Chapter 6 to estimate the fruit yield of an entire apple orchard. To do this, a continuous scan of the plantation will be carried out, analysing the videos recorded with the AK_VIDEO_ANALYSER software presented in Chapter 5. The resulting predictions will be contrasted with the actual harvest (fruit production) values.

In the medium term, even considering the development of fully automated fruit counting systems, their use should probably focus on sampling certain areas within an orchard, avoiding the consumption of excessive time and resources in generating and processing much more data than necessary. Sampling is therefore expected to remain a cornerstone in the process. Random sampling is, in theory, a very useful tool to avoid biasing the estimates (Wulfsohn, 2010). However, humans are poor at selecting individuals 'at random' and farmers do not have the confidence to sufficiently cover the orchard area, probably leaving gaps without taking data and then producing less precise estimates (Uribeetxebarria, et al., 2019a). To gain precision in the estimates (i.e., make lower errors), systematic sampling (or grid sampling) allows this problem to be overcome as the orchard area is more evenly covered (Webster and Lark, 2012). An interesting variant of systematic sampling is the so-called multilevel systematic sampling. As different studies have shown (Wulfsohn et al., 2012), systematic sampling first between trees and then on branches and shoots within the selected trees achieves improved precision using this multistage process compared to a simple random selection of trees. Fig. 7.1 tries to show the procedure of this type of nested and systematic sampling within a fruit tree.



(a) (b) (c)

Fig. 7.1 - A two-stage sampling design applied to a tree: (a) Stage 1. Primary branches are selected serving as primary sampling units (all fruits contained in branches 2, 4, 6, 8 and 10 are included in the sample after this first stage). Sampling period is set at m_1=2 (that is, using a sampling fraction $1/m_1$ =1/2), starting randomly the systematic selection at branch number 2. (b) Stage 2. Lateral branches (or branch segments) within primary branches are selected using also a sampling period m_2=2, but starting with sampling unit number 1. The final sample (filled circles) contains all fruit on branch segments 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 within the sampled branches. An unbiased estimate of the total number of fruit on the tree is obtained by multiplying the combined sampling period $m=m_1 \times m_2$ and fruits counted in the sample (adapted from Wulfsohn et al., 2012). (c) Half tree count instead of whole tree count when double the number of trees are sampled in random or stratified sampling processes (regulations in Spain; BOE, 2005).

More recently, stratified sampling has begun to be used on some farms (Miranda et al., 2019). Stratifying the sample normally consists of distributing the trees or sampling points in such a way that the different areas (strata) within the orchard with different expected yields are sampled. Ranked set sampling has also been proposed for this purpose. Uribeetxebarria et al. (2019b) obtained good results using this method in estimating peach fruit load using sample sizes of only 5 trees per plot. In the dilemma of which method to use, the so-called coefficient of error (*CE*) and its probability (confidence level) are two widely accepted statistics to aid decision making. Sampling methods certainly have limitations. For instance, reliable auxiliary information is necessary in stratified sampling, or the setting of appropriate sampling periods at each stage in the multilevel systematic sampling. However, it is expected that, by eliminating the manual counting of fruits, detection systems will contribute to more efficient samplings by both making it possible to sample a greater number of trees and reducing the human error that usually occurs in manual counting (Anderson et al., 2019).

Finally, obtaining predictive models of yield (combining fruit counting, size distribution and weight) for other fruit varieties beyond apples is another need for the sector. In short, intelligently merging fruit detection and sampling is the challenge we face in the coming years.

7.5 References

Aggelopoulou, K.D., Wulfsohn, D., Fountas, S., Gemtos, T.A., Nanos, G.D., Blackmore, S., 2010. Spatial variation in yield and quality in a small apple orchard. Precis. Agric. 11(5), 538–556. https://doi.org/10.1007/s11119-009-9146-9

Anderson, N.T., Underwood, J.P., Rahman, M.M., Robson, A., Walsh, K.B., 2019. Estimation of fruit load in mango orchards: tree sampling considerations and use of machine vision and satellite imagery. Precis. Agric. 20(4), 823–839. https://doi.org/10.1007/s11119-018-9614-1

BOE, 2005. Spanish Ministerial Order PRE/1950/2005. https://www.boe.es/eli/es/o/2005/06/17/pre1950/con> Accesed 7.7.21.

Faraway, J.J., 2016. Linear Models with R. Chapman and Hall/CRC, New York, NY. https://doi.org/10.1201/b17144

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., Amatya, S., Karkee, M., Zhang, Q., Lewis, K., 2015. Sensors and systems for fruit detection and localization: A review. Comput. Electron. Agric. 116, 8–19. https://doi.org/10.1016/j.compag.2015.05.021

Gregorio, E., Llorens, J., 2021. Sensing crop geometry and structure. In: Kerry, R., Escolà, A. (Eds.), Sensing Approaches for Precision Agriculture. Springer, Cham, pp. 59–92. https://doi.org/10.1007/978-3-030-78431-7_3.

Miranda, C., Santesteban, L.G., Urrestarazu, J., Loidi, M., Royo, J.B., 2018. Sampling stratification using aerial imagery to estimate fruit load in peach tree orchards. Agriculture 8(6).

https://doi.org/10.3390/agriculture8060078

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.

Plá, F., Juste, F., Ferri, F., & Vicens, M., 1993. Colour segmentation based on a light reflection model to locate citrus fruits for robotic harvesting. Comput. Electron. Agric. 9(1), 53–70. https://doi.org/10.1016/0168-1699(93)90029-Z

Uribeetxebarria, A., Martínez-Casasnovas, J.A., Escolà, A., Rosell-Polo, J.R., Arnó, J., 2019a. Stratified sampling in fruit orchards using cluster-based ancillary information maps: a comparative analysis to improve yield and quality estimates. Precision Agriculture 20(2), 179–192. https://doi.org/10.1007/s11119-018-9619-9

Uribeetxebarria, A., Martínez-Casasnovas, J.A., Tisseyre, B., Guillaume, S., Escolà, A., Rosell-Polo, J.R., Arnó, J., 2019b. 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

Webster, R., Lark, R.M., 2012. Field sampling for environmental science and management. In Field Sampling for Environmental Science and Management (1st ed.). Routledge. https://doi.org/10.4324/9780203128640

Wulfsohn, D., Aravena Zamora, F., Potin Téllez, C., Zamora Lagos, I., García-Fiñana, M., 2012. Multilevel systematic sampling to estimate total fruit number for yield forecasts. Precis. Agric. 13(2), 256–275. https://doi.org/10.1007/s11119-011-9245-2

Wulfsohn D., 2010. Sampling Techniques for Plants and Soil. Landbauforschung Völkenrode, Special Issue 340, 3–30.

Chapter 8: Conclusions

The main conclusions of the Doctoral Thesis are the following:

I. Concerning methods and challenges posed by fruit sizing using AI

I.1 Achieving reliable fruit sizing using RGB-D cameras and deep learning algorithms is an emerging issue in fruit research, and a challenge to face to consolidate modern, completely digital fruit growing. In the coming years, new advances in depth cameras, increasingly optimized algorithms, development of process-appropriate software, and greater availability of varied, extensive and higher quality training datasets are expected. Achieving fully operational commercial solutions will largely depend on success in these intermediate steps.

I.2 Facilitating fruit sizing to farmers and technical advisors through digital tools such as depth cameras is a strategic point for the fruit sector. The use of conventional manual fruit load estimation methods should be replaced by others that are less expensive and more reliable. The challenge is not easy and, furthermore, the end user is demanding. So efforts should be directed towards the use of sensors and AI tools that are easy to use, reliable, and with the possibility of fusing data from various sources of information. RGB-D camera like the Azure Kinect is the option that has been chosen given its good a priori specifications.

II. Concerning software for acquiring and processing data supplied by the Azure Kinect camera

II.1 AKFruitData is the open-source software that has been developed for the primary use of Azure Kinect cameras in orchard environments. The software includes two independent modules, AK_ACQS for data acquisition, and AK_FRAEX to allow images to be extracted from videos recorded with the camera for further analysis.

II.2 Using Python programming, the operation of AKFruitData has been satisfactorily tested including, i) simultaneous synchronized acquisition from several different cameras and sensors, ii) possibility of georeferencing data to obtain mappable spatial information, and iii) final output of RGB, depth, IR and point cloud data.

III. Concerning software for fruit detection and operation of sizing algorithms and yield prediction models using the Azure Kinect camera

III.1 AKFruitYield is the open-source software that, also designed in a modular way, allows the user to analyze images and videos recorded with the Azure Kinect camera and provided by the AKFruitData software. The AK_SW_BENCHMARKER tool makes it possible to apply and evaluate performances of different sizing algorithms and allometric yield prediction models on color and depth tree images containing previously detected and manually labeled apples. The second tool AK_VIDEO_ANALYSER makes it possible to perform sequential analysis of videos recorded at plot scale including automatic apple detection, size estimation and final apple yield prediction.

III.2 Like AKFruitData software, AKFruitYield has been designed under user-friendly criteria, having developed easy-to-use graphical interfaces aimed at both fruit growers and fruit consultants as well as end users working in research.

IV. Concerning benchmarking of sizing algorithms and allometric models for apple yield prediction

IV.1 The sizing algorithms to apply in non-occluded apples are varied. With a prediction error below 5% (MAPE), fitting rotated rectangles to the shape of apples is a good method to recommend.

IV.2 However, in non-occluded apples and due to a compensatory effect between sizing and allometry algorithms, weight prediction is more satisfactory adjusting circles, ellipses or even bounding boxes to the detected apples. The linear allometric model that uses the two axes of the apples as predictors does not exceed 5.1% error.

V. Final thoughts

V.1 Time-of-flight RGB-D cameras (like the Azure Kinect) offer a good option for sizing apples using computer vision algorithms for subsequent weight predictions made with appropriate allometric models. What's more, these are relatively inexpensive devices that perform well in agricultural environments under variable lighting conditions throughout the day.

V.2 Future works are planned to expand the current functionalities of AKFruitData and AKFruitYield, with the aim of consolidating the use of RGB-D cameras in fruit growing. Adjusting the performance of RGB-D cameras and processing software to new demands from the productive sector is therefore seen as key to extending digital yield prediction in fruit orchards.

V.3 The promising results of this thesis open up the possibility of using RGB-D cameras for real-time fruit orchard characterization in the short term. Efforts must now be directed at making available to the sector affordable tools with proven usefulness in improving decision-making and fruit management.

Chapter 9: List of contributions

9.1 *Journal papers included in the thesis*

- **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.
- **Miranda, J.C.**, Arnó, J., Gené-Mola, J., Fountas, S., Gregorio, E., 2023. AKFruitYield: Modular benchmarking and video analysis software for Azure Kinect cameras for fruit size and fruit yield estimation in apple orchards. SoftwareX 24, 101548. https://doi.org/10.1016/j.softx.2023.101548
- **Miranda, J.C.**, Arnó, J., Gené-Mola, J., Lordan, J., Asín, L., Gregorio, E., 2023. Assessing automatic data processing algorithms for RGB-D cameras to predict fruit size and weight in apples. Computers and Electronics in Agriculture 214, 108302. https://doi.org/10.1016/j.compag.2023.108302
- **Miranda, J.C.**, Gené-Mola, J., Zude-Sasse, M., Tsoulias, N., Escolà, A., Arnó, J., Rosell-Polo, J.R., Sanz-Cortiella, R., Martínez-Casasnovas, J.A., Gregorio, E., 2023. Fruit sizing using AI: A review of methods and challenges. Postharvest Biology and Technology 206, 112587. https://doi.org/10.1016/j.postharvbio.2023.112587

9.2 *Software in open-access repositories*

- **Miranda, J.C.**, Gené Mola, J., Arnó, J., Gregorio, E., 2022. AK_ACQS Azure Kinect Acquisition System. https://github.com/GRAP-UdL-AT/ak_acquisition_system/
- **Miranda, J.C.**, Gené Mola, J., Arnó, J., Gregorio, E., 2022. AKFruitData: AK_SM_RECORDER - Azure Kinect SM Recorder. https://github.com/GRAP-UdL-AT/ak_sm_recorder/
- **Miranda, J.C.**, Gené Mola, J., Arnó, J., Gregorio, E., 2022. AKFruitData: AK_FRAEX - Azure Kinect Frame Extractor. https://github.com/GRAP-UdL-AT/ak_frame_extractor/
- **Miranda, J.C.**, Arnó, J., Gené-Mola, J. Fountas, S., Gregorio, E. 2023. AKFruitYield: AK_SW_BENCHMARKER - Azure Kinect Size Estimation & Weight Prediction Benchmarker. https://github.com/GRAP-UdL-AT/ak_sw_benchmarker/
- **Miranda, J.C.**, Arnó, J., Gené-Mola, J. Fountas, S., Gregorio, E. 2023. AK_VIDEO_ANALYSER - Azure Kinect Video Analyser. https://github.com/GRAP-UdL-AT/ak_video_analyser/

9.3 *Scientific foreign-exchange*

- The PhD candidate has performed a four-months research stay (May-August 2023) in the Agricultural University of Athens (Greece) under the supervision of Dr. Spyros Fountas. During his stay, the doctoral student carried out fruit detection research applying deep learning techniques and completed a software article that resulted in Chapter 5 of this doctoral thesis. Additionally, he contributed by teaching programming fundamentals to group members and reviewing an article.

- Collaboration on tasks related to the thesis chapter "Methodology for 3D reconstruction and volume estimation of fruit trees by using low-cost RGB-D cameras". Research carried out by Iva Xhimitiku of the Centro di Ateneo di Studi e Attività Spaziali "Giuseppe Colombo" - CISAS, University of Padua - Italy. https://cisas.unipd.it/

9.4 Research project participations

- RTI2018-094222-B-I00. PAgFRUIT: Tecnologías de Agricultura de Precisión para optimizar el manejo del dosel foliar y la protección fitosanitaria sostenible en plantaciones frutales. Ministerio de Ciencia, Innovación y Universidades. 2019 – 2021.
- TED2021-131871B-I00. DIGIFRUIT - Sistemas de monitoreo de bajo coste en plantaciones frutales para Agricultura de Precisión basados en sensores fotónicos. Ministerio de Ciencia e Innovación / AEI. 2022 – 2024.
- PID2021-126648OB-I00. PAgPROTECT - Protección de cultivos de precisión para conseguir objetivos del Pacto Verde Europeo en uso eficiente y reducción de fitosanitarios mediante Agricultura de Precisión. Ministerio de Ciencia e Innovación / AEI / FEDER. 2022 – 2025
- 2021 LLAV 00088. Dispositiu basat en IA per a la detecció i mesura automàtica de fruits en camp (FruitsAIz). Ajuts per a projectes innovadors amb potencial d'incorporació al sector productiu (LLAVOR2021). AGAUR - Generalitat de Catalunya. 2022 – 2023.

This thesis focuses on the detection (counting) of fruits and estimation of their size and weight in apple orchards through the application of computer vision techniques. This work seeks to provide fruit growers with advanced tools and methodologies to help them make accurate harvest yield predictions. Counting (quantifying) and locating fruits represent previous steps to achieve these predictions. By knowing this information, fruit growers can schedule in advance the required resources for harvest and post-harvest (labor, transportation, storage), design sales strategies and, ultimately, optimize the profitability of their farms.

University of Lleida

March 2024