Computers and Electronics in Agriculture 177 (2020) 105687

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review

Longsheng Fu[a,c,d,e,*], Fangfang Gao[a], Jingzhu Wu[b], Rui Li[a], Manoj Karkee[e,*], Qin Zhang[e]

[a] College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
[b] Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China
[c] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China
[d] Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi 712100, China
[e] Center for Precision and Automated Agricultural Systems, Washington State University, Prosser, WA 99350, USA

ARTICLE INFO

Keywords:
Structured light
Time of flight
Active infrared stereo
Infrared image
Depth image

ABSTRACT

Fruit detection and localization are essential for future agronomic management of fruit crops such as yield prediction, yield mapping and automated harvesting. However, to perform robust and efficient fruit detection and localization in orchard is a challenging task under variable illumination, low-resolutions and heavy occlusion by neighboring fruits, foliage, or branches. Therefore, researches of fruit detection and localization by getting more information of objects are essential. RGB-D (Red, Green, Blue -Depth) cameras are promising sensors and widely used in fruit detection and localization given that they provide depth information and infrared information in addition to RGB information. After presenting a discussion on the advantages and disadvantages of RGB-D cameras with different depth measurement principles and application fields, this paper reviews various types of RGB-D sensor systems and image processing methods used for fruit detection and localization in the field. Finally, major challenges for the successful application of RGB-D camera-based machine vision system, and potential future directions for the research and development in this area are discussed.

1. Introduction

Fruit crop industries are under pressure of growing world population to increase production and quality while reducing the environmental impact in a sustainable way. This is a major challenge for agricultural communities, especially in a context of rising farming costs and a shortage of skilled labor (Gongal et al., 2015; Zhang et al., 2016). Efficient and sustainable agronomic management is required to reduce economic and environmental input costs while increasing orchard productivity (Bargoti and Underwood, 2017). Agricultural robots can take advantage of new sensing, artificial intelligence and computational technologies (among others) to respond to this challenge.

The use of robotics in orchards is increasing, particularly in yield prediction, yield mapping and automated harvesting. Yield prediction is made by manual counting of selected sample trees, leading to inaccurate predictions due to high variability in orchards (Gené-Mola et al., 2020). As for yield mapping, production maps provide useful information for fruit growers. Fruit orchards usually show spatial variability due to soil variations, fertility, water irrigation, among others (Häni et al., 2020). An analysis of yield maps helps farmers to determine the reasons for and find solutions to cope with this variability. Finally, hand harvesting is a hard and human-resource intensive labor. Automated harvesting with robots will meet the increasing labor demand to lower the human risk of injuries in orchards (Zhao et al., 2016; Zhang et al., 2018a). In all, the development of intelligent robots interacting with agricultural fields will increase the accuracy of tasks and reduce the consumption of resources without decreasing yield, making it a reasonable option for repeatable tasks.

Numerous research efforts have been reported on developing a machine vision system for target (e.g. fruits and branches in tree canopies) detection and localization, which is the first important step for achieving robotic operations in the orchard. A large number of these studies used RGB (Red, Green, Blue) cameras to detect and locate the targets (Li et al., 2015; Gongal et al., 2015; Zhao et al., 2016; Zhang et al., 2020a). RGB cameras provide affordable sensors, which allow fruits to be distinguished from other objects by fruit properties (color, geometric shape, and texture) or by using machine learning techniques (Fu et al., 2019; Sun et al., 2019; Zhang et al., 2020b). However, RGB

* Corresponding authors at: College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China (L. Fu); Center for Precision and Automated Agricultural Systems, Washington State University, Prosser, WA 99350, USA (M. Karkee).
E-mail addresses: fulsh@nwafu.edu.cn, longsheng.fu@wsu.edu (L. Fu), manoj.karkee@wsu.edu (M. Karkee).

https://doi.org/10.1016/j.compag.2020.105687
Received 13 July 2020; Received in revised form 30 July 2020; Accepted 31 July 2020
Available online 07 August 2020

Table 1
Comparison of sensor principles.

	Typical sensors	Advantages	Disadvantages
Structured Light	Kinect v1; Xtion PRO Live	High precision at close range measurement	Low precision at long range measurement; heavily affected by illumination and light reflection
Time of Flight	Kinect v2	High precision at long range measurement; less affected by illumination	Low image resolution; high power consumption; high hardware cost
Active Infrared Stereo	RealSense R200; RealSense D435	Small size; light weight; high image resolution	High computation cost; poor real-time performance

cameras are susceptible to field conditions and thus can affect detection and localization results. In addition, RGB cameras can only acquire two-dimensional (2D) information of the scene without the help of stereoscopic techniques.

Since the introduction of an affordable Kinect sensor by Microsoft (Redmond, WA, USA) in 2010, RGB-D (Red, Green, Blue -Depth) sensors have become an essential component of machine vision systems with various applications in agriculture including fruit detection and localization. The Kinect and similar sensors provide not only 2D information but also depth measurements and infrared information for each pixel in the images. The RGB-D sensing systems, therefore, inherently offer an opportunity to map the acquired depth information with RGB information to generate a high-density, textured three-dimensional (3D) point cloud of targets (Zhang et al., 2020a). Fruit detection can be achieved using only RGB images (Lehnert et al., 2017; Wang et al., 2017; Tu et al., 2018; Gené-Mola et al., 2019; Tu et al., 2020; Liu et al., 2020), depth images (Tu et al., 2018; Tu et al., 2020; Gené-Mola et al., 2019), or infrared images (Sa et al., 2016; Liu et al., 2020), or fusion of RGB and infrared images (Sa et al., 2016; Milella et al., 2019; Liu et al., 2020), RGB and depth images (Wang et al., 2012; Tu et al., 2018; Tu et al., 2020; Gené-Mola et al., 2019; Arad et al., 2020), or RGB and point cloud images (Nguyen et al., 2016; Sa et al., 2017; Tao and Zhou, 2017; Liu et al. 2018). The localization of the fruit can be achieved through depth information or point cloud information (Mai et al., 2015; Nguyen et al., 2016; Song et al., 2016; Tian et al., 2019; Zhang et al., 2019; Yang et al., 2019; Arad et al., 2020).

As there has been a rapid increase of RGB-D sensors and their applications in agriculture over the last decade, proper synthesis, and analysis of the literature in this area is instrumental in providing clear guidance on the state-of-the-art and potential future direction of the RGB-D sensors, which is lacking in the literature so far. Therefore, this review is focused on providing up-to-date information on studies carried out so far in the area of fruit detection and localization in specialty crops using RGB-D sensors, as well as discussing potential challenges and opportunities of this technique. The remainder of this paper is structured as follows: Section 2 introduces and compares commonly used consumer RGB-D sensors. Sections 3 and 4 summarize the studies conducted using RGB-D sensors to detect and locate fruits, respectively. Challenges in the implementation of fruit detection and localization techniques for fruit harvesting as well as future trends of research and development in this area are discussed in Section 5 whereas a comprehensive summary of the paper is presented in Section 6.

2. Consumer RGB-D sensors

RGB-D sensors generally operate using one of the three depth measurement principles: i) structured light (SL), ii) time of flight (ToF), and iii) active infrared stereo (AIRS) technique. Sensors working with SL emit a light pattern onto the scene and calculate the depth information based on the deformation of the pattern. In most cases, the light emitted would be in the infrared spectrum and is thus invisible to the human eye. ToF sensors, on the other hand, use an infrared light emitter to emit a pulse of light and calculate the distance based on the time the signal takes to travel to and return back from the target objects to the detector and the speed of light. The third category of sensors operates with AIRS, which combines the idea of an active projection of light (e.g. structured light) and a passive stereo camera pair (where natural light is used to collect stereo pair images). In contrast to classical stereo cameras, active stereo cameras additionally project their own texture. Thus, they can gather information, even on low-textured surfaces.

The sensors are divided into active and passive systems depending on if separate illumination source is required to operate these sensors. And all three kinds of sensors are active systems because they utilize infrared light projection (750–1400 nm) to obtain depth data (Kuan et al., 2019). Given that sunlight covers the entire spectrum of infrared light, the performance of all sensors may suffer when used outdoors. The sensors operating with SL technique are vulnerable to ambient illumination and multi-device interference, whereas ToF systems suffer from motion artifacts and multi-path interference. The AIRS-based sensors, on the other hand, suffer from common stereo matching issues such as occluded pixels, which can cause over smoothing, edge fattening, and flying pixels near contour edges. Besides, there are also some other differences between the three sensors, and the comparison of them is shown in Table 1.

2.1. Cameras based on structured light

As mentioned before, a SL sensor is a scanning device for measuring the 3D shape of an object using projected light patterns. The first examples of SL systems appeared in computer vision literature in the 1990s (Boyer and Kak, 1987; Chen et al., 1997) and have been widely investigated since. The first consumer-grade SL depth camera products only hit the mass market in 2010 with the introduction of the first-generation Kinect (Microsoft, Redmond, WA, USA) based on Primesensor (PrimeSense, Tel Aviv, Israel). The Primesensor also appeared in other consumer products, such as Xtion PRO Live sensor (Asus, Taipei, Taiwan) and the Structure Sensor (Occipital, Boulder, CO, USA). Subsequently, other SL depth cameras reached the market, such as the RealSense F200 and SR300 (Intel, Santa Clara, CA, USA).

The first-generation Kinect sensor (Kinect v1) is one of the most prominent sensors operating with SL and has been well embraced by, e.g., the robotics community. Also driven by its low price, Kinect v1 has been widely applied in agricultural applications, such as fruit detection (Dionisio et al., 2016; Nguyen et al., 2016; Méndez Perez et al., 2017), fruit localization (Wang et al., 2012, 2016), orchard management (Xiao et al., 2017), livestock monitor (Salau et al., 2017), and crop phenotyping (Paulus et al., 2014; Wang and Li, 2014; Nguyen et al., 2015; Yamamoto et al., 2015; Dionisio et al., 2016). A sensor with specifications and capabilities/features comparable to Kinect v1 is Xtion PRO Live. This sensor has also been used in various agricultural applications (García-Luna and Morales-Díaz, 2016; Li et al., 2019; Lao et al., 2019). However, with no specified performance/error value, slightly higher price, and marginally lower resolution, it offers no distinctive advantage over Kinect v1 given the current application area.

Similar to Microsoft's efforts, Intel's RealSense cameras F200 and SR300 were released to target the assessment of 3D point clouds at short distances between 0.2 and 2 m. Their most salient advantages include a substantially lower price compared to all other cameras considered in this paper, and higher framerate of 60 fps. Therefore,

RealSense F200 and SR300 were specially employed in the near range agricultural applications, such as fruit detection and localization (Lehnert et al., 2017; Liu et al., 2017, 2018; Ramos et al., 2018; Vit and Shani, 2018; Milella et al., 2019).

2.2. Time of flight cameras

The ToF sensor is a range imaging camera system that calculates distance based on the known speed of light, and the estimated time-of-flight of a light signal between the camera and the subject at each point in the captured images. The depth map resolution of ToF cameras generally does not reach video graphic array resolution of 640 × 480 pixels, which is much lower than that of SL cameras, but their frame rate is generally higher than that of SL cameras. ToF sensors operate with full irradiation of emitting light, while SL irradiates to only local areas, so the power consumption of ToF is higher. One of the most important features of the ToF sensors is that the depth calculation is not affected by the gray level and other features of the surface of the object leading to generally more accurate 3D measurement. Besides, the depth calculation accuracy of ToF does not change with distance, which is very important for maintaining the performance in application with large-scale motions.

One of the most widely used ToF cameras is the second-generation of Kinect (Kinect v2) (Microsoft, Redmond, WA, USA). Compared to the first generation, because of the limitation of the principle of ToF, this camera has a comparatively lower resolution of depth sensor (512 × 424 instead of 640 × 480) but a slightly extended range of admissible depth values (4.5 instead of 3.5 m). Its most important advantage is the increased horizontal and vertical viewing angle, allowing obtaining more overlapping regions in consecutive depth images. As shown in Fig. 1, Kinect v1 and v2 were placed in the same position to capture images in a 'Scifresh' apple orchard (Prosser, WA, USA) at the same time. All the images (RGB, depth, and infrared) obtained with Kinect v2 were of better quality than those with Kinect v1. Therefore, Kinect v2 has been employed more widely in agricultural applications, such as fruit detection (Choi et al., 2015; Kusumam et al., 2017; Pamornnak et al., 2017; Tao and Zhou, 2017; Wang et al., 2017; Tu et al., 2018; Zhang et al., 2018b; Tu et al., 2020; Liu et al., 2020; Kang and Chen, 2020b; Lin et al., 2020), fruit localization (Mai et al., 2015; Kusumam et al., 2017; Tian et al., 2019; Yang et al., 2019; Zhang et al., 2019), livestock management (Kongsro, 2014; Misimi et al., 2016; Guo et al., 2017; Nir et al., 2018; Yin et al., 2019), and crop phenotyping (Dionisio et al., 2016; Rosell-Polo et al., 2017; Haemmerle and Hoefle, 2018; Vázquez-Arellano et al., 2018; Ma et al., 2019; Arad et al., 2020).

There are also some ToF based depth sensors, such as CamBoard pico flex/monstor (PMD Technologies GmbH, Siegen, NRW, German) and Swissranger SR4000/4500 (Heptagon, Zurich, Zurich, Switzerland). Especially the Swissranger SR4500 is the most expensive depth sensor that costs around 4,000 euro (Munaro et al., 2016). It offers a quite different depth sensing ability to measure distances between 0.8 and 9 m with an estimation error below four centimeters. Those sensors were used with RGB cameras to obtain RGB and depth information simultaneously for crop recognition (Li and Tang, 2018) and fruit detection (Fernández et al., 2014; Barnea et al., 2016; Gongal et al., 2018).

2.3. Active infrared stereo cameras

The AIRS cameras are an extension of the traditional passive stereo vision method. They are using an infrared stereo camera pair with a pseudorandom pattern projective texturing the scene via a patterned infrared light source to cope with low-texture environments. The infrared light emitted by the camera also gives it the ability to capture images at night. With a proper selection of the signal wavelength, the camera pair captures a combination of active illumination and passive light, improving quality above that of SL while providing a robust solution in both indoor and outdoor scenarios. Although this technology was introduced decades ago, it has only recently become available in commercial products (e.g., Intel RealSense R200 and D400 family). Recently released model D435 (Intel, Santa Clara, CA, USA) provides a much higher resolution of 1920 × 1080 in RGB image and 1280 × 720 in the depth map. It can capture four images (RGB, depth, left infrared, right infrared) simultaneously, as shown in Fig. 2, which was obtained in the same field as Fig. 1.

There are some problems specific to the AIRS-based camera. They must process very high-resolution images to match the high-frequency patterns that produced from their projector. It required that many local minima should be avoided from alternating arrangement of these high-frequency patterns. Besides, the camera must compensate for differences in brightness between the projected patterns on the surface. Additionally, this camera cannot train with supervising on a large active stereo dataset with true ground depth since no data is available. Because these sensors have been available for only a short period, there is relatively little prior work utilizing AIRS-based depth imaging utilized in agriculture. Two example studies found in the literature were Milella et al. (2019) that used RealSense R200 for field grapevine phenotyping and Kang and Chen (2020b) that used RealSense D435 to detect apples in the orchard.

2.4. Comparison of various consumer RGB-D cameras

There are many different types of consumer RGB-D cameras operating on various 3D measurement principles, and they all have their own characteristic. All the RGB-D cameras can transfer data by USB 2/3. Some of them are also powered by USB 2/3, such as all the RealSense cameras, Structure Sensor, and Xtion PRO Live. However, Kinect v1 and v2 require AC/DC power supply, which limits their outdoor applications. In terms of weight, the RealSense D415/435 cameras weight only 72 g, which offers a potential to be installed in a wide range of sensing platforms. An empirical evaluation of ten most common depth sensors in terms of bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments were reported by Georg et al. (2019).

Another important attribute of an RGB-D sensor is the field of view (FoV) of the depth sensor, which can be used to compare and contrast all sensors/cameras as depicted in Table 2. Software Development Kits (SDKs) and the supported programming languages (PLs) for accessing the sensor software Application Programing Interfaces (APIs) were also considered as important factors for comparing the sensors and their areas of application include. Finally, sensor price was considered as an essential, non-technical attribute for our comparison matrix, which can be used to calculate a price-performance ratio.

3. RGB-D sensors for fruit detection

Detection results of machine vision systems in detecting fruit in tree canopies is affected by uncertain and variable illumination conditions in the field, variable, and complex canopy structures and varying color, shape, and size of the fruit. Also, it is substantially limited by the occlusion of fruit in canopy images by leaves, branches, and other fruits. Several studies have been carried out to accurately detect fruit in outdoor orchard environment (Tang et al., 2020). Methods studied in the past for fruit detection utilized different types of sensors and investigated various conventional and soft computational methods for image analysis (Wang et al., 2020). Also, some studies have focused on classifications based on explicitly extracted features of target objects (e.g. fruit) to improve the detection rate potentially. In recent years, however, deep learning-based methods (often end-to-end processing without pre-processing and explicit feature extraction) have been more widely used in fruit detection. Various types of RGB-D sensors and image segmentation methods used for fruit detection are reviewed in the following sub-sections and summarized in Table 3.

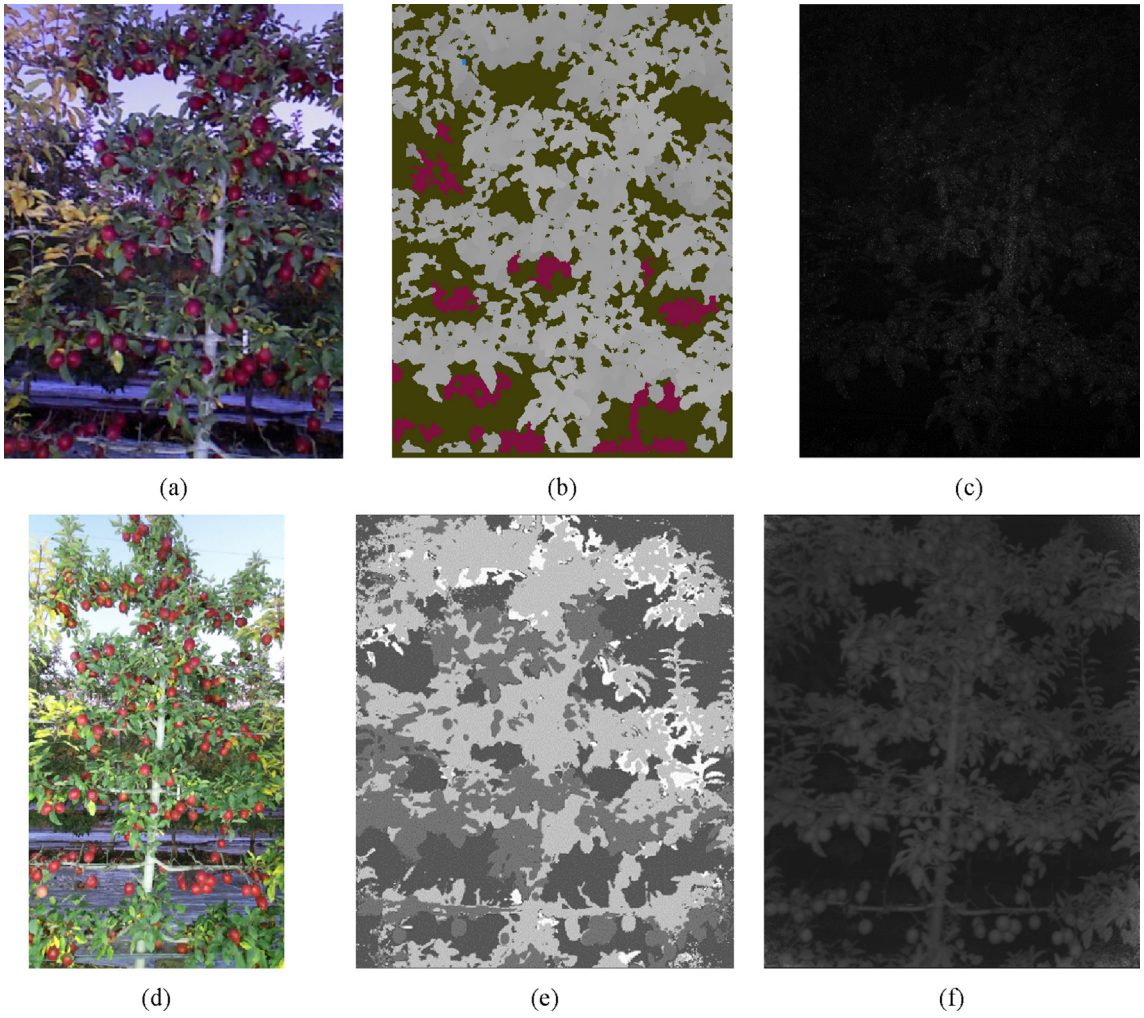

Fig. 1. Outdoor images captured with Kinect v1 and v2 at the same location and same time in a commercial 'Scifresh' apple orchard (Prosser, WA, USA). The top row includes RGB image with 640 × 480 pixels (a), depth image with 640 × 480 pixels (b), and infrared image with 640 × 480 pixels (c) from Kinect v1 whereas the bottom row includes RGB image with 1920 × 1080 pixels (d), depth image with 512 × 424 pixels (e), and infrared image with 512 × 424 pixels (f) from Kinect v2.

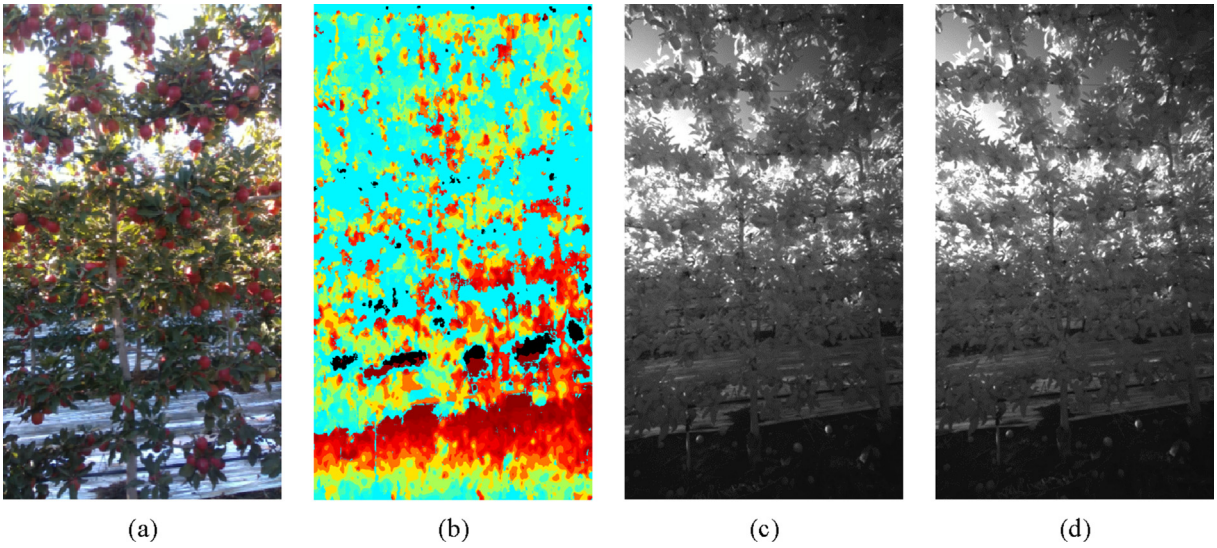

Fig. 2. Example images captured concurrently with RealSense D435 camera in a commercial 'Scifresh' apple orchard (Prosser, WA, USA). (a) RGB image with 1920 × 1080 pixels; (b) Depth image with 1280 × 720 pixels; (c) Left infrared image with 1280 × 720 pixels; (d) Right infrared image with 1280 × 720 pixels.

Table 2
Comparison of various consumer RGB-D cameras.

Camera	Principle	Measuring range / m	Error / mm	Res. RGB	Res. Depth	Frame rate / fps	Depth FoV (H°×V°)	Interface	PL or SDK	Price / USD
Structure Sensor	SL	0.4–3.5	1%*	640 × 480	640 × 480	30/60	58 × 45	USB2/ Lighting	C/C++	379
Kinect v1	SL	0.8–3.5	< 400	640 × 480	640 × 480	15/30	57 × 43	USB 2	C#/C++/VB/JAVA/Matlab	150
Xtion PRO Live	SL	0.8–3.5	–	1280 × 1024	640 × 480	30/60	58 × 45	USB 2	C#/C++/JAVA	199
RealSense F200	SL	0.2–1.2	1%*	1920 × 1080	640 × 480	60	73 × 59	USB 3	C#/C++/JAVA/JavaScript	99
RealSense SR300	SL	0.3–2.0	1%*	1920 × 1080	640 × 480	30/60	55 × 71.55	USB 3	C#/C++/JAVA/JavaScript	129
Senz3D	ToF	0.2–1.0	–	1080 × 720	320 × 240	30	74 × 41.6	USB 2	C++/C	199
Kinect v2	ToF	0.5–4.5	< 1%*	1920 × 1080	512 × 424	15/30	70.6 × 60	USB 3	C#/C++/C/Matlab	200
CamBoard Pico Flex	ToF	0.1–4.0	< 6	n/a	224 × 171	45	82 × 66	USB 2	C++/C/Matlab	389
CamBoard Pico Monstar	ToF	0.5–6.0	< 1%*	n/a	352 × 287	60	100 × 85	USB 3	C++/C/Matlab	1935
RealSense ZR300	AIRS	0.55–2.8	–	1920 × 1080	628 × 468	30/60	68 × 41	USB 3	C#/C++/JAVA/JavaScript	630
RealSense R200	AIRS	0.5–4.0	–	1920 × 1080	640 × 480	30/60	59 × 46	USB 3	C#/C++/JAVA/JavaScript/Python	99
RealSense D415	AIRS	0.16–10.0	< 2%	1920 × 1080	1280 × 720	30/60/90	63 × 40	USB 3	C#/C++/JAVA/JavaScript/Python	149
RealSense D435	AIRS	0.2–10.0	< 2%	1920 × 1080	1280 × 720	30/60/90	91.2 × 65.5	USB 3	C#/C++/JAVA/JavaScript/Python	179
RealSense D435i	AIRS	0.11– 10.0	–	1920 × 1080	1280 × 720	30/90	85 × 58	USB 3	C#/C++/JAVA/JavaScript/Python	199

Note. *, of the measured distance; -, not specified; Res., resolution; USB, universal serial bus; PL, Programming Language; SDK, Software Development Kit.

3.1. Features and methods for fruit detection

3.1.1. Fruit detection with RGB images

Color, geometric shape, and texture information in RGB image are important features used in the machine vision system to distinguish fruit from leaves, branches, and other background objects in the orchard. Many studies have used RGB image-based segmentation/classification techniques to detect fruits with different features. Wang et al. (2017) employed Kinect v2 to recognize on-tree mangoes using the histogram of oriented gradients and an ellipse-fitting algorithm based on the RGB images and achieved a detection rate of 81.1%. Tu et al. (2018) used RGB images of passion fruit obtained by Kinect v2 as the input of a Faster RCNN with VGG16, and its detection rate reached 89.7%. Tu et al. (2020) applied RGB images obtained by Kinect v2 to detect passion fruit using multiple scale faster region-based convolutional neural networks (MS-FRCNN), and an F1-score of 91.6% was obtained. Gené-Mola et al. (2019) and Liu et al. (2020) also used the original RGB images acquired by Kinect v2 as the input to Faster R-CNN with VGG16 to detect apples (80.8%) and kiwifruits (88.4%), respectively. As can be seen from the results, these approaches achieved a detection rate of more than 80% when only RGB images were used for fruit detection. However, the fruit detection results based on RGB image is affected by variation in fruit features due to its maturity level, crop variety, uncertain and varying background features, and variable lighting conditions. For example, depending upon the maturity level and fruit exposure to the sun over the growing season, color of 'Scifresh' apples in a tree canopy may vary from green, reddish, reddish-green to red, which makes it challenging to develop fruit detection methods that work robustly on all those apples. Besides, fruit under direct sunlight and shadow during imaging will have different light intensity in the images, and thus would be challenging to detect them with only the RGB information.

3.1.2. Fruit detection with depth images

The depth image refers to an image that takes the distance from the sensor to each point in the scene as a pixel value. The sensor uses a black and white spectrum to sense the distance from itself to the point in the scene. The RGB-D camera collects every point in the FoV and forms a depth of field image representing the surrounding environment (Tu et al., 2018). In the obtained depth image, the deeper the black is, the closer the target is; the lighter the white is, the farther the target is. The gray area between the black and white represents the physical distance between the target and the sensor. Therefore, the depth image was uniquely used for detecting fruit. Gené-Mola et al. (2019) obtained depth image by Kinect v2 to detect apples using Faster RCNN with VGG16, but its average precision (AP) was only 61.3%. Tu et al. (2018) used RGB images of passion fruit obtained by Kinect v2 as the input of Faster RCNN with VGG16, and its detection rate was 76.3%. Tu et al. (2020) applied depth images obtained by Kinect v2 to detect passion fruit using MS-FRCNN, and an F1-score of 84.1% was obtained. The detection rate based on depth images is lower than that based on RGB images from the same camera with the same FoV, 19.5% lower in Gené-Mola et al. (2019), 13.4% lower in Tu et al. (2018), and 7.5% lower in Tu et al. (2020). The depth images are more susceptible to the external environment than RGB images, which causing much noises in the depth images and thus affect detection results. Therefore, to achieve fruit detection more accurate and efficient, information other than depth information should not be ignored.

3.1.3. Fruit detection with infrared images

Infrared sensor is another distinguishable feature of the RGB-D cameras. It detects infrared energy (temperature) and converts it into an electronic signal to form an image, which is helpful in differentiating fruit and background in tree canopies. Fruit, generally, absorbs more heat and radiates more heat compared to leaves and other parts of the plant canopies, which allows for the distinction between those plant

Table 3
Summary of different features and methods for RGB-D sensors used in fruit detection.

Features	Sensors	Detection rate	Speed / s	Segmentation method	Crops applied	Reference
RGB image	Kinect v2	81.1%	–	Histogram of oriented gradients	Mango	Wang et al. (2017)
	Kinect v2	89.7%	0.072	Faster R-CNN with VGG16	Passion fruit	Tu et al. (2018)
	Kinect v2	88.7%	–	Faster R-CNN with VGG16	Apple	Gené-Mola et al. (2019)
	Kinect v2	91.6%	0.175	MS-FRCNN	Passion fruit	Tu et al. (2020)
	Kinect v2	88.4%	0.134	Faster R-CNN with VGG16	Kiwifruit	Liu et al. (2020)
Depth image	Kinect v2	76.3%	0.050	Faster R-CNN with VGG16	Passion fruit	Tu et al. (2018)
	Kinect v2	61.3%	–	Faster R-CNN with VGG16	Apple	Gené-Mola et al. (2019)
	Kinect v2	84.1%	0.145	MS-FRCNN	Passion fruit	Tu et al. (2020)
Infrared image	Kinect v2	79.7%	–	Faster R-CNN with VGG16	Sweet pepper	Sa et al. (2016)
	Kinect v2	89.2%	0.135	Faster R-CNN with VGG16	Kiwifruit	Liu et al. (2020)
Integration of RGB and infrared image	Kinect v2	83.8%	–	Faster R-CNN with VGG16	Sweet pepper	Sa et al. (2016)
	RealSense R200	91.5%	–	Faster R-CNN with VGG19	Grape	Milella et al. (2019)
	Kinect v2	91.7%	0.134	Faster R-CNN with VGG16	Kiwifruit	Liu et al. (2020)
Integration of RGB and depth image	Kinect v1	–	–	Color thresholding algorithm	Apple	Wang et al. (2012)
	Kinect v2	92.7%	0.072	Faster R-CNN with VGG16	Passion fruit	Tu et al. (2018)
	Kinect v2	88.0%	–	Faster R-CNN with VGG16	Apple	Gené-Mola et al. (2019)
	Kinect v2	94.6%	0.175	MS-FRCNN	Passion fruit	Tu et al. (2020)
	Fotonic F80	85.0%	–	SSD network with VGG16	Sweet pepper	Arad et al. (2020)
Integration of RGB and point cloud	Kinect v2	91.0%	< 1.000	Color threshold of R-G	Apple	Mai et al. (2015)
	Kinect v1	91.0%	0.964	Circular Hough transformation and random sample consensus algorithm	Apple	Nguyen et al. (2016)
	Kinect v2	92.3%	–	SVM based on Genetic algorithm	Apple	Tao and Zhou (2017)
	RealSense SR300	71%	–	Point Feature Histograms	Sweet pepper	Sa et al. (2017)
	RealSense SR300	–	–	Naïve Bayes classifier	Sweet pepper	Lehnert et al. (2017)
	RealSense F200	63.8% to 100%	–	Depth-sphere cutting	Six varieties of oranges	Liu et al. (2018)

Note: -, not available.

materials using infrared imaging (Stajnko et al., 2004; Wachs et al., 2010). Sa et al. (2016) input the RGB image (acquired by Kinect v2) and infrared image (acquired by JAI multi-spectral camera) to Faster RCNN with VGG16 for detecting sweet pepper. The F1-score of RGB image (0.816) is higher than that of infrared image (0.797). However, these two kinds of images were acquired by two cameras, it is hard to say that the results of using infrared image to detect crops are inferior to those using RGB image.

The results of using different images acquired by the same camera as the detected images in the study are persuasive. Liu et al. (2020) used the infrared images of kiwifruit captured by Kinect v2 as the input to a Faster R-CNN with VGG16 and achieved an AP of 89.2%. In this study, the results were compared with aligned RGB images that acquired by the same Kinect v2, which showed a slight improvement in AP value (by 0.8%) from RGB image to infrared image. The primary reason is that the infrared image has a larger intensity gap between the object boundary and other pixels, which makes the edge detection process more accessible. Because the intensity difference between the border pixels and other pixels is small, edge detection in RGB images is relatively more difficult than infrared images.

3.1.4. Integration of RGB and infrared images

Variable illuminations in orchards affect the intensity of reflected light, resulting in uneven exposure on the acquired image, which has been affecting the detection performance. On the other hand, occlusion caused by leaves, branches, and clusters of fruits will affect the geometric features of fruits in the images. In such scenarios, detection based on a single image (e.g., RGB, depth, or infrared) may not be the best approach. Multi-modality sensors offered by the RGB-D cameras may be more beneficial because different sensors can provide complementary information regarding various aspects of the fruit and background.

By registering and fusing the information on RGB images and infrared images acquired with RGB-D cameras, the detection rate of fruits can be improved to a certain extent. Sa et al. (2016) combined the RGB and infrared information for detecting sweet pepper through Kinect v2, which improved the F1-score from 79.7% to 83.8% compared to single-modal (infrared) information. Milella et al. (2019) generated a dense 3D map of the grapevine row based on infrared stereo reconstruction and augmented with its color appearance, which detected fruit by a RealSense R200 camera with a detection rate of 91.5%. Besides, the kiwifruit image obtained by the feature fusion of the RGB image and infrared image captured by the Kinect v2 was input into the Faster R-CNN with VGG16, and the AP of 90.7% was reported (Liu et al., 2020).

3.1.5. Integration of RGB and depth images

In some studies, RGB images and depth images obtained with RGB-D cameras have also fused to generate a point cloud for fruit detection. Wang et al. (2012) used the color, shape, and depth features obtained with Kinect v1 to detect on-tree apples in an indoor environment and reported that all ten target fruits were successfully detected. However, this algorithm/study was not conducted on an outdoor apple orchard that has more fruits with real complicated appearances. Gené-Mola et al. (2019) obtained an AP of 88.0% by integrating depth and RGB images of apple orchards acquired by Kinect v2. Similarly, Arad et al. (2020) used Fotonic F80 camera that simultaneously acquired RGB and depth information to detect sweet peppers, and achieved a detection rate of 85% using SSD with VGG16 under the most suitable crop conditions. Tu et al. (2018) used Kinect v2 to acquire passion fruit images in the orchard, and obtained a detection rate of 92.7% when integrated RGB and depth images were used, which was 3% higher than the detection rate of only RGB images used and 16.4% higher than the detection rate of only depth images used. Similarly to the study of detecting passion fruit by Kinect v2 in Tu et al. (2020), the F1-score increased from 91.6% of using only RGB images to 94.6% of integrating RGB and depth images, and a 10.5% increase when using integrated RGB and depth images comparing with only depth images. Tu et al. (2020) concluded that a detection system that combines depth and RGB detection/information could improve the detection results under different illumination conditions.

3.1.6. Integration of RGB and point cloud

Depth images such as those acquired with RGB-D cameras can be converted to a point cloud format with x, y, and z coordinates. Integration of point clouds and RGB information for fruit detection is also a common method used for fruit detection. Mai et al. (2015) used R-G color threshold segmentation method to segment the point cloud of apple fruit and background acquired by Kinect v2, and achieved a fruit detection rate of 91.0% on average. Nguyen et al. (2016) exploited both color and shape properties to recognize the apples in point cloud obtained by Kinect v1. Results showed that 100% of the fully visible apples and 82% of the partially occluded apples were detected correctly. Lehnert et al. (2017) developed effective vision algorithms based on color and 3D point cloud data acquired by RealSense SR300 to enable successful harvesting of sweet peppers. Tao and Zhou (2017) extracted an improved 3D descriptor with the fusion of color features and 3D geometry features from the pre-processed point clouds acquired by Kinect v2 to detect apple with a detection rate of 92.3%. Sa et al. (2017) used color and geometry information acquired from RealSense F200 sensor to detect sweet pepper and obtained an AP of 71%. Liu et al. (2018) utilized point cloud data in a close range of 160 mm from RealSense F200 to estimate different geometric features of the fruit and leaf to recognize citrus fruits. The fruit detection rates ranged from 80% to 100% when occlusion and clustering of fruit were minimal. However, heavier occlusion and clustering of fruit still caused a great influence on the overall success rate of fruit recognition in tree canopies leading to a significantly lower detection rate of 63.8%.

3.2. Image segmentation methods for fruit detection

3.2.1. Support vector machine

Support vector machine (SVM) is a supervised statistical learning algorithm that has been used for linear and non-linear regression analysis and pattern recognition/classification. For linearly separable classification, SVM separates the two classes with a maximum margin between them by a hyper-linear plane. Similarly, in non-linearly separable classification, feature vectors are mapped to a new feature space that is linearly separable, and then linear SVM separation is used to classify features in the new space.

SVM has been widely used to classify images for agricultural applications. Barnea et al. (2016) exploited both RGB and depth data obtained by a SwissRanger depth camera to analyze shape-related features of sweet peppers both in the image plane and 3D space, and followed classification of the resultant feature vector using an SVM classifier. Sa et al. (2017) served a concatenated feature vector that build by both color and geometry information acquired from a RealSense SR300 as the input for SVM to detect the sweet pepper peduncle, and the study obtained an AP of 71%. Tao and Zhou (2017) used SVM to detect apples, branches, and leaves in an image obtained by Kinect v2, and achieved a detection rate of 94.64%, 47.05%, and 75.00%, respectively. Lin et al. (2019) trained a color, gradient, and geometry feature-based SVM using Kinect v2 RGB-D images to remove false positives in citrus detection, which was robust with an F1-score of 91.97%. Wu et al. (2020) fused color and 3D geometry features of peach using a SVM classifier to segment the point cloud obtained by Kinect v2, which had a fruit detection rate of 80.1%.

3.2.2. Artificial neural network

An artificial neural network (ANN) is a supervised learning algorithm that can learn from the data through an iterative training process as it improves its performance after running each iteration up to a certain degree. ANNs consists of computational nodes or neurons arranged in multiple layers and interconnections between them that feed the information from one part of the network to others. Strength or weight of these connections adopted iteratively during the training process to learn specific patterns/model defined by the training data and store the learned knowledge. The model created this way can then be used to transform inputs provided at one end of the network to the computed results that are available at the other end of the network. The number of layers and neurons in each layer of the network depend on the complexity of the system being modelled (Dahikar and Rode, 2014). Arefi and Motlagh (2013) developed an expert system based on wavelet transform and ANN for detecting ripe tomatoes. A total of 90 wavelet features were extracted from each tomato, and a feed-forward neural network was used to distinguish the ripe tomato from its background. A detection rate of 95.5% was obtained with the proposed recognition algorithm. In an effort by Kurtulmus et al. (2011), a statistical classifier, an ANN, and an SVM classifier were built and used for detecting peaches. The study reported that successful fruit detection rates of 84.6%, 77.9%, and 71.2% were achieved respectively with the three classifiers listed above when they were tested with the same test dataset. However, none of the fruit images used in these studies was obtained using RGB-D cameras. More recently, remarkable progress has been achieved through the introduction of deep learning, which is based on multiple layer artificial neural networks (Koirala et al., 2019). Therefore, there are many studies based on deep learning to use the images obtained by RGB-D cameras for fruit detection.

3.2.3. Deep learning

Deep learning belongs to the machine learning computational field and is similar to ANN. However, deep learning is about “deeper” neural networks that provide a hierarchical representation of the data by means of various convolutions. This allows larger learning capabilities and thus higher performance and precision. Deep learning-based object detection is gaining popularity in recent years, which has been applied in many different research fields (Gao et al., 2020). Deep neural networks such as convolutional neural network (CNN), region-based CNN (R-CNN), Fast R-CNN, Faster R-CNN, You Only Look Once (YOLO) network and their improvements (Bargoti and Underwood, 2017; Fu et al., 2018; Gené-Mola et al., 2019) provide an excellent framework for fruit detection as well.

Inspired by potentials for high accuracy, reliability and robustness, and real-time computation, fruit detection based on deep learning techniques have also been studied widely on images from the RGB-D sensors. Tu et al. (2018) detected passion fruits through Faster R-CNN and achieved a detection rate of 92.7% with Kinect v2. Tu et al. (2020) improved architecture of multiple scales Faster R-CNN by incorporating feature maps from shallower convolution feature maps for regions of interest pooling and obtained an F1-score of 94.6% for passion fruit with Kinect v2. Sa et al. (2016) used Faster R-CNN with VGG16 to detect sweet peppers and achieved an F1-score of 83.8%. Similarly, Gené-Mola et al. (2019) and Liu et al. (2020) also employed Faster R-CNN with VGG16 to detect apples (94.8% of AP) and kiwifruit (90.7% of AP) by Kinect v2, respectively, whereas Milella et al. (2019) applied the same model to detect grape clusters with a detection rate of 91.5% by RealSense R200. Fu et al. (2020) used Faster with VGG16 to detect Foreground-RGB apple images obtained by depth information of Kinect v2, and a AP of 89.3% was achieved. Zhang et al. (2020c) deployed transfer learning and fine-tuning for Faster R-CNN with VGG19 and activated the feature of different layers to detect apples, branches, and trunks with images obtained by Kinect v2, thus achieving the highest mean AP of 82.4%. Yang et al. (2019) added three maximum pooling layers to the YOLO-V3 convolutional layer module to enhance the model ability to extract citrus features, and achieved an F1-score of 85.0% with Kinect v2. Kang and Chen (2020a) proposed a deep neural network DaSNet-v2 with resnet-101 to detect apple RGB images from RealSense D435 and obtained 87.3% of F1-score. Besides, Kang and Chen (2020b) also proposed a deep-learning-based fruit detector LedNet, which reached an F1-score of 84.9% on apple image from Kinect v2 in the orchard. These studies, in general, showed that remarkable progress has been achieved in fruit detection in field conditions through the application of deep learning.

Table 4
Summary of features and methods for fruit localization.

Features	Sensors	Localization error / mm	Speed / s	Main method	Crop applied	Reference
Depth Information	Fotonic F80	–	3.7	SSD network with VGG16	Sweet pepper	Arad et al. (2020)
	Kinect v2	5.9	0.4	Improved YOLO v3	Citrus	Yang et al. (2019)
	RealSense R200	7.0	–	Color thresholding algorithm	Strawberry	Xiong et al. (2019)
	Kinect depth camera	1.9	–	Image segmentation based on the color and region growing	greenhouse cucumber	Song et al., (2016)
	Kinect v2	–	–	Image segmentation,	Apple	Tian et al. (2019)
Point Cloud Information	Kinect v2	8.1	< 1	Random sample consensus	Apple	Mai et al. (2015)
	Kinect v2	–	–	Faster RCNN	Apple	Zhang et al. (2019)
	Kinect v1	< 10	–	Circular Hough transformation and random sample consensus algorithm	Apple	Nguyen et al. (2016)

4. Features and methods for fruit localization

Fruit localization in crop canopies is another essential part of the machine vision system for robotic working. Numerous studies have been carried out in the past several decades to locate fruit in tree canopies with reasonable successes (Harrell et al., 1990; Slaughter and Harrell, 1989; Mehta and Burks, 2014; Font et al., 2014). However, RGB-D cameras have been applied to fruit localization systems only in recent years (Mai et al., 2015; Nguyen et al., 2016; Song et al., 2016; Xiong et al., 2019; Tian et al., 2019; Zhang et al., 2019; Yang et al., 2019; Arad et al., 2020). The following sections and Table 4 will provide an overview of the various types of RGB-D sensors and methods used for fruit localization.

4.1. Localization with depth information

As mentioned before, depth information obtained with different kinds of depth sensors embedded with RGB-D cameras has been used in recent years for fruit localization in the field. In general, the 2D coordinate information of the fruit is obtained through the detection result, whereas the depth information is acquired from the depth image obtained by RGB-D camera. Then, some specific coordinate transformation methodologies has been adapted to convert the 2D coordinates of the center point of the target into 3D coordinates to achieve the localization of the fruit. Arad et al. (2020) employed the depth information extracted by Fotonic F80 depth camera from the detected sweet pepper regions and used a standard procedure of pixel-to-world transformation of the region to calculate 3D location of the point of mass. The average fruit localization time of this technique was 3.7 s. Yang et al. (2019) used Kinect v2 to convert 2D center point coordinates of citrus into 3D coordinates with a localization error of 5.9 mm. Similarly, Xiong et al. (2019) combined the depth information and intrinsic matrix provided by RealSense R200 for calculating 3D location of strawberry and achieved a localization error of 7.0 mm.

In other studies, coordinates of the center of fruits were estimated without explicitly implementing any fruit detection techniques. Song et al. (2016) segmented the RGB image of greenhouse cucumber based on the color and region growing to extract the feature values of the tangent point and the centroid. Then the 3D coordinates of the feature point of the fruit were determined by fusing the depth information obtained by a Kinect depth camera and image feature point coordinates. The localization error achieved was about 1.9 mm, which meets the requirements of a picking robot. In another study, Tian et al. (2019) obtained gradient information from the depth image obtained by Kinect v2. Then a vortex was formed using the gradient information, and the vortex center was determined as the estimated center of the corresponding apple. Both of these studies acquired the target center coordinates based on depth image without fruit detection and achieved promising results, providing a new technique for fruit detection and localization.

4.2. Localization with point cloud information

Instead of depth information, point clouds obtained with RGB-D cameras can also be used for fruit localization. Some studies focused on using point cloud of the entire scene generated by fusing the RGB information and depth information. In this case, a point cloud of fruit regions will be acquired by segmenting the whole point cloud. Zhang et al. (2019) collected point cloud images of apple trees by Kinect v2 using this method, and then segmented the fruit regions by the point cloud segmentation method of the interest region in the RGB image. The proposed method achieved segmented purity of 96.7% and 96.2% for red and green apples, respectively. Mai et al. (2015) obtained the 3D position information and radius of fruits by segmented the point cloud acquired by Kinect v2 using a color threshold, thereby acquiring the centroid of each fruit. Average fruit positioning error with this technique was 8.1 mm, and the average error in estimating fruit radius was 4.5 mm. Nguyen et al. (2016) pre-processed point clouds obtained by Kinect v1 using distance filter and color filter, and segmented the point cloud using a Euclidean distance-based clustering algorithm. The position and diameter of each apple are estimated by separating and combining the clusters, which achieved a position estimation error of < 10 mm.

5. Challenges and future trends

As discussed before, most of the studies in fruit detection and localization suggested that occlusions, clustering, and variable lighting conditions in the field environment posed major challenges limiting fruit detection as well as localization accuracies. Further research and development would be essential to address these challenges and improve accuracy, reliability and robustness of image segmentation, fruit detection and localization systems with low-cost RGB-D cameras. These advancements are necessary to establish the technical as well as the commercial viability of RGB-D camera-based fruit detection and localization systems.

5.1. Hardware improvement

5.1.1. Improving camera performance

Improving various features/attributes of the RGB-D cameras such as durability and resolution can help improve fruit detection and localization accuracies. Operating principle of depth sensors in RGB-D cameras (e.g. SL, ToF, AIRS) inherently leads to certain advantages and disadvantages of the sensors, therefore warranting specific types of improvement that might be possible with each of the cameras for optimal performance in fruit detection and localization.

The sensor technology based on the principle of SL is relatively mature. It offers relatively higher resolution, and simpler calculation, and is suitable for application in indoor (Nguyen et al., 2015). However, the speckle infrared spot of the SL core technology will be submerged under strong ambient light, so the depth information obtained based on

this principle can be greatly affected by strong sunlight (Kuan et al., 2019). Therefore, improving the RGB-D camera allows it to acquire and emit the weakest waveband of the infrared spectrum of sunlight in a certain period to obtain fruit images, which can minimize the impact of sunlight on image quality. However, the weakest wavelength band of the solar spectrum within a certain period of time cannot be obtained in the current technical, so the above solution cannot be realized for the time being. Moreover, the infrared-emitting device can be damaged comparatively quickly with sustained use, which affects detection and localization results. Therefore, the performance of these cameras could be enhanced by improving their infrared emitting devices.

As discussed before, depth cameras operating based on the principle of ToF use infrared emitters to emit light pulses and calculate the distance based on the time period until the detector receives the pulse back from the target and the speed of light. ToF depth camera can change the measured distance of the camera by adjusting the frequency of the transmitted pulse to calculate the depth value. The depth information obtained by the camera is not affected by the grayscale and characteristics of the subject, so its measurement accuracy will not decrease with the increase of the measurement distance (Zanuttigh et al., 2016). However, depth cameras based on this principle are also affected by ambient light. Compared to RGB-D cameras operating with other principles, these cameras are more expensive, and the resolution of the obtained images is lower. Therefore, improving the camera resolution and reducing the cost could be an important area for future research and development on these cameras. ToF technology is now being frequently applied to the rear lens of mobile phones, providing these devices a capability to have precise depth perception. Therefore, the target and background can be accurately identified in the photograph, thereby creating a professional-level background blur effect (Fan et al., 2018). Improving the RGB-D camera so that it also can blur the background and retain complete target information could be beneficial to fruit detection and localization research in certain situation of working with a modern fruiting wall orchard.

The depth cameras operating with the principle of AIRS rely on pure image feature matching, so the quality of the acquired images will be inferior when the light is dim or overexposed. However, compared with the classical stereo cameras, the active stereo cameras can also project its texture, thus making up for the lack of feature extraction and matching in the case of the lack of texture in the scene under study (Milella et al., 2019). If the intensity of the infrared emitted by the AIRS cameras is higher than that of the ambient light, the influence of ambient light, including sunlight, can be minimized by combining with an optical filter of the corresponding wave band. Although the current RGB-D sensor technologies have not yet realized the concept, this could be a good direction for future research and development.

5.1.2. Improving illumination conditions

Adjusting the illuminations could also be a possible approach for reducing the influence of uncertain and variably lighting and improving fruit detection and localization accuracies. In some cases, such as the scorching sun in summer, the light intensity can be tremendous, which can cause overexposure of the photosensitive sensor arrays. Therefore, some studies tried to build a shading platform to eliminate the effects of ambient illumination. Nguyen et al. (2016) constructed a light shield consisted of black plastic foil to block direct sunlight. Besides, a tunnel structure was created by covering an imaging platform with tarpaulin on four sides and the top of the platform (Gongal et al., 2016). And a controlled, uniform lighting environment was created by installing a number of Light Emitting diode (LED) in the platform (Gongal et al., 2016; Silwal et al., 2016; Gongal et al., 2018).

Although the shielding device covering the entire tree is large, which greatly reduces the efficiency of the overall system for fruit detection and localization but the shielding device indeed reduces the effects of ambient illumination. Therefore, a portable shielding device can be designed to save the time of building and dismantling each time images are acquired, thereby improving the efficiency of fruit detection and localization. In the absence of a shielding device, a filter can be used to filter the strong sunlight and the LED to supplement the weak sunlight can offset the unstable sunlight to a certain extent, improving the quality of the obtained fruit images.

5.2. Algorithm improvement

There is always a room for improvement in methodologies and algorithms used to detect and localize fruit, which can make a varying level of impact in the end results obtained depending on the nature and complexity of the application at hand. Some basic image pre-processing techniques such as histogram equalization, noise filtering, and removal of surface shadows on the acquired images, can enhance the quality of the image. More information can be obtained in these processed images, thereby reducing the impact of the changes in illumination on the image (Lv et al., 2019). In recent years, deep learning-based techniques have shown to be more robust and less affected by variable lighting conditions than traditional methods. Combining the depth information acquired by RGB-D cameras with the RGB and infrared information has also resulted in improved robustness in minimizing the impact of lighting condition and improve fruit detection results (Milella et al., 2019). Liu et al. (2020) used RGB and infrared images acquired by Kinect v2 as the input to the Faster R-CNN with VGG16 and obtained the AP of 88.4% and 89.2%, respectively. When the two images were fused as input, the AP increased slightly to 90.7%. In the future, the fusion of more information, including RGB information, infrared information, and depth information, can be studied to see whether it can improve the detection and localization accuracies.

The further improved algorithm allows robots to perform multiple tasks simultaneously, which can improve the efficiency of the robot. Kang and Chen (2020) proposed an improved deep neural network DaSNet-v2, which can perform detection and instance segmentation on fruits and semantic segmentation on branches to provide rich information for each object, especially for those overlapping fruits. The results of this research include not only the detected fruits but also the distribution of the branches in fruit tree canopies. Such a system, in the future, can be used to devise control strategies to guide automated robot harvesting in a more efficient and collision-free manner. Algorithms improvement enable robotic machines to concurrently perform multiple tasks can be a promising direction for future researches.

5.3. Human-machine collaboration

Human-computer/machine collaboration can make up for the deficiencies in all functions of a machine vision system to a certain extent. The commercial application of fruit robot is still unavailable because of a lack of high efficiency and economic justification (Zhao et al., 2016). One of the approaches to improve the applicability of robotic harvesting (and other robotic applications in agriculture) is to combine human workers and robotic systems synergistically (Ge et al., 2019; Arad et al., 2020). Existing fruit detection systems always have undetectable fruits because of various limitations discussed before. However, if some level of support is provided by human operators for detecting fruit and stem position and orientation in complex areas where fruits are heavily occluded and correct the errors made by the machine, overall speed and detection rate of the system may potentially be improved. Besides, the operator can also manually label fruits that were not detected on the vision system, so that the fruit harvesting robot has the opportunity to be more efficient and also learn from the examples to be better in the future.

5.4. Modifying crop canopies for improved fruit visibility and accessibility

A well-structured orchard, specifically trellis trained systems with

narrow canopies (e.g. fruiting wall orchards), is the basis for minimizing fruit occlusion and achieving high level of fruit detection and localization accuracies, and to facilitate the operation of robot machines such as harvesters. In terms of managing orchards, a systematic method of various horticultural operations, including tree training, pruning, pollination, and flower and green fruit (or fruitlet) thinning have the potential to present fruit to the machine vision system with high visibility and few occlusion and clustering (Gongal et al., 2015). Cultivating fruit trees with desired growth habits suitable for vision systems and automated/robotic operations in combination with horticultural technology mentioned above will fundamentally improve the overall performance of the machine vision systems and followed operations such as robotic harvesting. For example, modern planting model in apple and kiwifruit orchards make the fruit appear drooping, which not only reduces the ratio of fruit overlapping but also greatly facilitates the image processing and robotic picking operations. The narrow canopy designs of fruit crops need to be adopted in wider crop growing regions around the world and many kinds of crops such as apples, pears, peaches and citrus.

6. Summary

Over the past decade, RGB-D sensors have been widely used in agricultural applications such as fruit detection, fruit localization, orchard management, livestock monitors, and crop phenotyping. This paper provided a comprehensive review of fruit detection and localization studies using RGB-D cameras. The advantages and disadvantages of various types of RGB-D cameras with different depth measurement principles and application fields were also discussed. The cameras operating with the principle of SL commonly used for fruit detection and positioning/localization include Kinect v1 and RealSense F200 and SR300. One of the most important benefits of these sensors is their low cost. The commonly utilized ToF-based sensor for fruit detection and localization is Kinect v2, which has been used more widely than SL-based sensors. However, this camera suffers from comparatively low resolutions of depth and infrared images. Recently, RGB-D cameras operating with the principle of AIRS have also been commercially released. Though these cameras haven't been widely applied in agriculture so far, they have shown to be promising for fruit detection and localization based on the ongoing research at Washington State University and other institutions around the world.

Following the introduction of various types of RGB-D sensors, this review discussed various image processing methods used for fruit detection with the images acquired with these sensors. Fruit detection can be achieved by using RGB images, depth images/information, infrared images, or fusion of two or more types of images/information. Some studies reported that the fruit detection rate achieved with infrared images was higher than the same with RGB images, whereas RGB image-based detection was found to be better than the same with depth images. Besides, the fruit detection results with the fusion of multiple types of information were found to be better than the same with only one type of information.

This study also discussed various types of RGB-D sensor systems and image processing methods for fruit localization. The localization of the fruit can be achieved using either the depth images or the point cloud information. Irrespective of the type of depth information used to locate fruits, the co-registered RGB, depth, and infrared information acquired by RGB-D cameras plays a crucial role for improved detection performance and robustness of these techniques. However, compared with the volume of research on fruit detection with RGB-D sensing systems, there are fewer studies on the localization of fruits with these sensors, and therefore has a wider need and scope for future research and development.

Finally, the current challenges and potential further researches in fruit detection and location were discussed in the paper. Most of the past studies suggested that occlusions, clustering, and variable lighting conditions were the major challenges for the accurate detection and localization of fruit in the field. Thus, to improve fruit detection and localization performance for robotic harvesting and other applications, further researches should focus on the ways to address these limitations and improve the detection rate, robustness, and speed of the sensing and image processing systems while reducing the overall complexity and cost. Improvements in camera performance and algorithms, sensing platform that can improve the uniformity of lighting condition, horticultural modifications, and human–machine collaboration could be some of the areas for future research.

CRediT authorship contribution statement

Longsheng Fu: Conceptualization, Methodology, Supervision, Writing - original draft, Writing - review & editing. **Fangfang Gao:** Conceptualization, Investigation, Methodology, Writing - original draft. **Jingzhu Wu:** Methodology, Supervision, Writing - review & editing. **Rui Li:** Investigation, Supervision, Writing - review & editing. **Manoj Karkee:** Conceptualization, Methodology, Writing - review & editing. **Qin Zhang:** Methodology, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

The authors express their deep gratitude to the Young Faculty Study Abroad Program of the Northwest A&F University Scholarship who sponsored Dr Longsheng Fu in conducting post-doctoral research at the Centre for Precision and Automated Agricultural Systems, Washington State University, and to the Allan Brothers Fruit Company who provided the experimental orchard.

Funding

This work was supported by the China Postdoctoral Science Foundation funded project (2019M663832); Fundamental Research Funds for the Central Universities of China (2452020170); Key Research and Development Program in Shaanxi Province of China (grant number 2018TSCXL-NY-05-04, 2019ZDLNY02-04); National Natural Science Foundation of China (grant number 31971805); International Scientific and Technological Cooperation Foundation of Northwest A&F University (grant number A213021803).

References

Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellström, T., Hemming, J., Kurtser, P., Ringdahl, O., Tielen, T., van Tuijl, B., 2020. Development of a sweet pepper harvesting robot. J. F. Robot. 1–13. https://doi.org/10.1002/rob.21937.

Arefi, A., Motlagh, A.M., 2013. Development of an expert system based on wavelet transform and artificial neural networks for the ripe tomato harvesting robot. Aust. J. Crop Sci. 7, 699–705.

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

Barnea, E., Mairon, R., Ben-Shahar, O., 2016. Colour-agnostic shape-based 3D fruit detection for crop harvesting robots. Biosyst. Eng. 146, 57–70. https://doi.org/10.1016/j.biosystemseng.2016.01.013.

Boyer, K.L., Kak, A.C., 1987. Color-encoded structured light for rapid active ranging. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-9, 14–28. https://doi.org/10.1109/TPAMI.1987.4767869.

Chen, C., Hung, Y., Chiang, C., Wu, J., 1997. Range data acquisition using color structured lighting and stereo vision. Image Vis. Comput. 15, 445–456. https://doi.org/10.1016/s0262-8856(96)01148-1.

Choi, D., Lee, W.S., Ehsani, R., Roka, F.M., 2015. A machine vision system for quantification of citrus fruit dropped on the ground under the canopy. Trans. ASABE 58, 933–946. https://doi.org/10.13031/trans.58.10688.

Dahikar, S.S., Rode, S.V., 2014. Agricultural crop yield prediction using artificial neural network approach. Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. 2, 683–686. https://doi.org/10.1016/j.indcrop.2018.09.055.
Dionisio, A., Angela, R., César, F.-Q., José, D., 2016. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. Comput. Electron. Agric. 122, 67–73. https://doi.org/10.1016/j.compag.2016.01.018.
Fan, Y., Feng, Z., Mannan, A., Khan, T.U., Shen, C., Saeed, S., 2018. Estimating tree position, diameter at breast height, and tree height in real-time using a mobile phone with RGB-D SLAM. Remote Sens. 10, 1845. https://doi.org/10.3390/rs10111845.
Fernández, R., Salinas, C., Montes, H., Sarria, J., 2014. Multisensory system for fruit harvesting robots. Experimental testing in natural scenarios and with different kinds of crops. Sensors 14, 23885–23904. https://doi.org/10.3390/s141223885.
Font, D., Pallejà, T., Tresanchez, M., Runcan, D., Moreno, J., Martínez, D., Teixidó, M., Palacín, J., 2014. A proposal for automatic fruit harvesting by combining a low cost stereovision camera and a robotic arm. Sensors 14, 11557–11579. https://doi.org/10.3390/s140711557.
Fu, L., Feng, Y., Majeed, Y., Zhang, X., Zhang, J., Karkee, M., Zhang, Q., 2018. Kiwifruit detection in field images using Faster R-CNN with ZFNet. IFAC-PapersOnLine 51, 45–50. https://doi.org/10.1016/j.ifacol.2018.08.059.
Fu, L., Majeed, Y., Zhang, X., Karkee, M., Zhang, Q., 2020. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. Biosyst. Eng. 197, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.07.007.
Fu, L., Tola, E., Al-Mallahi, A., Li, R., Cui, Y., 2019. A novel image processing algorithm to separate linearly clustered kiwifruits. Biosyst. Eng. 183, 184–195. https://doi.org/10.1016/j.biosystemseng.2019.04.024.
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.
García-Luna, F., Morales-Díaz, A., 2016. Towards an artificial vision-robotic system for tomato identification. IFAC-PapersOnLine 49, 365–370. https://doi.org/10.1016/j.ifacol.2016.10.067.
Ge, Y., Xiong, Y., Tenorio, G.L., From, P.J., 2019. Fruit localization and environment perception for strawberry harvesting robots. IEEE Access 7, 147642–147652. https://doi.org/10.1109/ACCESS.2019.2946369.
Gené-Mola, J., Gregorio, E., Auat Cheein, F., Guevara, J., Llorens, J., Sanz-Cortiella, R., Escolà, A., Rosell-Polo, J.R., Escolà, A., 2020. 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., Vilaplana, V., Rosell-Polo, J.R., Morros, J.R., Ruiz-Hidalgo, J., Gregorio, E., 2019. 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.
Georg, H.F., Markus, S., Martin, K., Markus, V., 2019. An empirical evaluation of ten depth cameras: Bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments. IEEE Robot. Autom. Mag. 26, 67–77. https://doi.org/10.1109/MRA.2018.2852795.
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, 498–503. https://doi.org/10.1016/j.inpa.2018.06.002.
Gongal, A., Silwal, A., Amatya, S., Karkee, M., Zhang, Q., Lewis, K., 2016. Apple crop-load estimation with over-the-row machine vision system. Comput. Electron. Agric. 120, 26–35. https://doi.org/10.1016/j.compag.2015.10.022.
Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., Zhu, D., 2017. LSSA_CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs. Comput. Electron. Agric. 138, 60–68. https://doi.org/10.1016/j.compag.2017.04.014.
Haemmerle, M., Hoefle, B., 2018. Mobile low-cost 3D camera maize crop height measurements under field conditions. Precis. Agric. 19, 630–647. https://doi.org/10.1007/s11119-017-9544-3.
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.
Harrell, R.C., Adsit, P.D., Munilla, R.D., Slaughter, D.C., 1990. Robotic picking of citrus. Robotica 8, 269–278. https://doi.org/10.1017/S0263574700000308.
Kang, H., Chen, C., 2020a. Fast implementation of real-time fruit detection in apple orchards using deep learning. Comput. Electron. Agric. 168, 105108. https://doi.org/10.1016/j.compag.2019.105108.
Kang, H., Chen, C., 2020b. 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.
Koirala, A., Walsh, K.B., Wang, Z., McCarthy, C.L., 2019. 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.
Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008.
Kuan, Y.weng, Ee, N.O., Wei, L.S., 2019. Comparative study of intel R200, Kinect v2, and primesense RGB-D sensors performance outdoors. IEEE Sens. J. 19, 8741–8750. https://doi.org/10.1109/JSEN.2019.2920976.
Kurtulmus, F., Lee, W.S., Vardar, A., 2011. Green citrus detection using “eigenfruit”, color and circular Gabor texture features under natural outdoor conditions. Comput. Electron. Agric. 78, 140–149. https://doi.org/10.1016/j.compag.2011.07.001.
Kusumam, K., Krajník, T., Pearson, S., Duckett, T., Cielniak, G., 2017. 3D-vision based detection, localization, and sizing of broccoli heads in the field. J. F. Robot. 34, 1505–1518. https://doi.org/10.1002/rob.21726.
Lao, C., Yang, H., Li, P., Feng, Y., 2019. 3D reconstruction of maize plants based on consumer depth camera. Trans. Chinese Soc. Agric. Mach. 50, 222–228. https://doi.org/10.6041/j.issn.1000-1298.2019.07.024.
Lehnert, C., English, A., McCool, C., Tow, A.W., Perez, T., 2017. Autonomous sweet pepper harvesting for protected cropping systems. IEEE Robot. Autom. Lett. 2, 872–879. https://doi.org/10.1109/LRA.2017.2655622.
Li, J., Huang, W., Zhao, C., 2015. Machine vision technology for detecting the external defects of fruits - A review. Imaging Sci. J. 63, 241–251. https://doi.org/10.1179/1743131X14Y.0000000088.
Li, J., Tang, L., 2018. Crop recognition under weedy conditions based on 3D imaging for robotic weed control. J. F. Robot. 35, 596–611. https://doi.org/10.1002/rob.21763.
Li, P., Lao, C., Yang, H., Feng, Y., 2019. Maize plant 3D information acquisition system based on mobile robot platform. Trans. Chinese Soc. Agric. Mach. 50, 15–21. https://doi.org/10.6041/j.issn.1000-1298.2019.S0.003.
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, 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, J., Yuan, Y., Zhou, Y., Zhu, X., Syed, T.N., 2018. Experiments and analysis of close-shot identification of on-branch citrus fruit with realsense. Sensors 18, 1510. https://doi.org/10.3390/s18051510.
Liu, J., Zhu, X., Yuan, Y., 2017. Depth-sphere transversal method for on-branch citrus fruit recognition. Trans. Chinese Soc. Agric. Mach. 48, 32–39. https://doi.org/10.6041/j.issn.1000-1298.2017.10.004.
Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., Cui, Y., 2020. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access 8, 2327–2336. https://doi.org/10.1109/ACCESS.2019.2962513.
Lv, J., Wang, Y., Xu, L., Gu, Y., Zou, L., Yang, B., Ma, Z., 2019. A method to obtain the near-large fruit from apple image in orchard for single-arm apple harvesting robot. Sci. Hortic. 257, 108758. https://doi.org/10.1016/j.scienta.2019.108758.
Ma, X., Feng, J., Zhang, W., Guan, H., Zhu, K., 2019. Calculation method for maize plant height based on depth information. Int. Agric. Eng. J. 28, 325–332.
Mai, C., Zheng, L., Sun, H., Yang, W., 2015. Research on 3D reconstruction of fruit tree and fruit recognition and location method based on RGB-D camera. Trans. Chinese Soc. Agric. Mach. 46, 35–40. https://doi.org/10.6041/j.issn.1000-1298.2015.S0.006.
Mehta, S.S., Burks, T.F., 2014. Vision-based control of robotic manipulator for citrus harvesting. Comput. Electron. Agric. 102, 146–158. https://doi.org/10.1016/j.compag.2014.01.003.
Méndez Perez, R., Cheein, F.A., Rosell-Polo, J.R., 2017. Flexible system of multiple RGB-D sensors for measuring and classifying fruits in agri-food Industry. Comput. Electron. Agric. 139, 231–242. https://doi.org/10.1016/j.compag.2017.05.014.
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.
Misimi, E., Øye, E.R., Eilertsen, A., Mathiassen, J.R., Åsebø, O.B., Gjerstad, T., Buljo, J., Skotheim, Ø., 2016. GRIBBOT - Robotic 3D vision-guided harvesting of chicken fillets. Comput. Electron. Agric. 121, 84–100. https://doi.org/10.1016/j.compag.2015.11.021.
Munaro, M., Basso, F., Menegatti, E., 2016. OpenPTrack: Open source multi-camera calibration and people tracking for RGB-D camera networks. Rob. Auton. Syst. 75, 525–538. https://doi.org/10.1016/j.robot.2015.10.004.
Nguyen, T.T., Slaughter, D.C., Max, N., Maloof, J.N., Sinha, N., 2015. Structured light-based 3D reconstruction system for plants. Sensors 15, 18587–18612. https://doi.org/10.3390/s150818587.
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.
Nir, O., Parmet, Y., Werner, D., Adin, G., Halachmi, I., 2018. 3D Computer-vision system for automatically estimating heifer height and body mass. Biosyst. Eng. 173, 4–10. https://doi.org/10.1016/j.biosystemseng.2017.11.014.
Pamornnak, B., Limsiroratana, S., Khaorapapong, T., Chongcheawchamnan, M., Ruckelshausen, A., 2017. An automatic and rapid system for grading palm bunch using a Kinect camera. Comput. Electron. Agric. 143, 227–237. https://doi.org/10.1016/j.compag.2017.10.020.
Paulus, S., Behmann, J., Mahlein, A.K., Plümer, L., Kuhlmann, H., 2014. Low-cost 3D systems: Suitable tools for plant phenotyping. Sensors 14, 3001–3018. https://doi.org/10.3390/s140203001.
Ramos, P.J., Avendaño, J., Prieto, F.A., 2018. Measurement of the ripening rate on coffee branches by using 3D images in outdoor environments. Comput. Ind. 99, 83–95. https://doi.org/10.1016/j.compind.2018.03.024.
Rosell-Polo, J.R., Gregorio, E., Gene, J., Llorens, J., Torrent, X., Arno, J., Escola, A., 2017. Kinect v2 sensor-based mobile terrestrial laser scanner for agricultural outdoor applications. IEEE/ASME Trans. Mechatronics 22, 2420–2427. https://doi.org/10.1109/TMECH.2017.2663436.
Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., McCool, C., 2016. Deepfruits: A fruit detection system using deep neural networks. Sensors 16, 1222. https://doi.org/10.3390/s16081222.
Sa, I., Lehnert, C., English, A., McCool, C., Dayoub, F., Upcroft, B., Perez, T., 2017. Peduncle detection of sweet pepper for autonomous crop harvesting-combined color

and 3-D information. IEEE Robot. Autom. Lett. 2, 765–772. https://doi.org/10.1109/LRA.2017.2651952.
Salau, J., Haas, J.H., Junge, W., Thaller, G., 2017. Automated calculation of udder depth and rear leg angle in Holstein-Friesian cows using a multi-Kinect cow scanning system. Biosyst. Eng. 160, 154–169. https://doi.org/10.1016/j.biosystemseng.2017.06.006.
Silwal, A., Karkee, M., Zhang, Q., 2016. A hierarchical approach to apple identification for robotic harvesting. Trans. ASABE 59, 1079–1086. https://doi.org/10.13031/trans.59.11619.
Slaughter, D.C., Harrell, R.C., 1989. Discriminating fruit for robotic harvest using color in natural outdoor scenes. Trans. Am. Soc. Agric. Eng. 32, 757–763. https://doi.org/10.13031/2013.31066.
Song, J., Teng, D., Wang, K., 2016. Segmentation and localization method of greenhouse cucumber based on image fusion technology. Int. J. Simul. Syst. Sci. Technol. 17, 11–14. https://doi.org/10.5013/IJSSST.a.17.25.07.
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, 31–42. https://doi.org/10.1016/S0168-1699(03)00086-3.
Sun, S., Li, C., Paterson, A.H., Chee, P.W., Robertson, J.S., 2019. Image processing algorithms for infield single cotton boll counting and yield prediction. Comput. Electron. Agric. 166, 104976. https://doi.org/10.1016/j.compag.2019.104976.
Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., Zou, X., 2020. Recognition and localization methods for vision-based fruit picking robots : A review. Front. Plant Sci. 11, 510. https://doi.org/10.3389/fpls.2020.00510.
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.
Tian, Y., Duan, H., Luo, R., Zhang, Y., Jia, W., Lian, J., Zheng, Y., Ruan, C., Li, C., 2019. Fast recognition and location of target fruit based on depth information. IEEE Access 7, 170553–170563. https://doi.org/10.1109/ACCESS.2019.2955566.
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. https://doi.org/10.1007/s11119-020-09709-3.
Tu, S., Xue, Y., Zheng, C., Qi, Y., Wan, H., Mao, L., 2018. Detection of passion fruits and maturity classification using Red-Green-Blue Depth images. Biosyst. Eng. 175, 156–167. https://doi.org/10.1016/j.biosystemseng.2018.09.004.
Vázquez-Arellano, M., Reiser, D., Paraforos, D.S., Garrido-Izard, M., Burce, M.E.C., Griepentrog, H.W., 2018. 3-D reconstruction of maize plants using a time-of-flight camera. Comput. Electron. Agric. 145, 235–247. https://doi.org/10.1016/j.compag.2018.01.002.
Vit, A., Shani, G., 2018. Comparing RGB-D sensors for close range outdoor agricultural phenotyping. Sensors 18, 4413. https://doi.org/10.3390/s18124413.
Wachs, J.P., Stern, H.I., Burks, T., Alchanatis, V., 2010. Low and high-level visual feature-based apple detection from multi-modal images. Precis. Agric. 11, 717–735. https://doi.org/10.1007/s11119-010-9198-x.
Wang, B., Chen, Z., Gao, J., Fu, L., Su, B., Cui, Y., 2016. The acquisition of kiwifruit feature point coordinates based on the spatial coordinates of image. IFIP Adv. Inf. Commun. Technol. 478, 399–411. https://doi.org/10.1007/978-3-319-48357-3_39.
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., Mao, W., Liu, G., Hu, X., Li, S., 2012. Identification and location system of multi-operation apple robot based on vision combination. Trans. Chinese Soc. Agric. Mach. 43, 165–170. https://doi.org/10.6041/j.issn.1000-1298.2012.12.030.
Wang, W., Li, C., 2014. Size estimation of sweet onions using consumer-grade RGB-depth sensor. J. Food Eng. 142, 153–162. https://doi.org/10.1016/j.jfoodeng.2014.06.019.
Wang, Z., Walsh, K.B., Verma, B., 2017. On-tree mango fruit size estimation using RGB-D images. Sensors 17, 2738. https://doi.org/10.3390/s17122738.
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 accuracy. Comput. Electron. Agric. 174, 105475. https://doi.org/10.1016/j.compag.2020.105475.
Xiao, K., Ma, Y., Gao, G., 2017. An intelligent precision orchard pesticide spray technique based on the depth-of-field extraction algorithm. Comput. Electron. Agric. 133, 30–36. https://doi.org/10.1016/j.compag.2016.12.002.
Xiong, Y., Peng, C., Grimstad, L., From, P.J., Isler, V., 2019. Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. Comput. Electron. Agric. 157, 392–402. https://doi.org/10.1016/j.compag.2019.01.009.
Yamamoto, S., Hayashi, S., Tsubota, S., 2015. Growth measurement of a community of strawberries using three-dimensional sensor. Environ. Control Biol. 53, 49–53. https://doi.org/10.2525/ecb.53.49.
Yang, C., Liu, Y., Wang, Y., Xiong, L., Xu, H., Zhao, W., 2019. Research and experiment on recognition and location system for citrus picking robot in natural environment. Trans. Chinese Soc. Agric. Mach. 50 (14–22), 72. https://doi.org/10.6041/j.issn.1000-1298.2019.12.002.
Yin, L., Cai, G., Tian, X., Sun, A., Shi, S., Zhong, H., Liang, S., 2019. Three dimensional point cloud reconstruction and body size measurement of pigs based on multi-view depth camera. Trans. Chinese Soc. Agric. Eng. 35, 201–208. https://doi.org/10.11975/j.issn.1002-6819.2019.23.025.
Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., Cortelazzo, G.M., 2016. Operating principles of time-of-flight depth cameras. Time-of-Flight and Structured Light Depth Cameras: Technol. App. 81–133. https://doi.org/10.1007/978-3-319-30973-6.
Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., Gao, Z., 2018a. Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN). Comput. Electron. Agric. 155, 386–393. https://doi.org/10.1016/j.compag.2018.10.029.
Zhang, J., Karkee, M., Zhang, Q., Zhang, X., Yaqoob, M., Fu, L., Wang, S., 2020a. Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting. Comput. Electron. Agric. 173, 105384. https://doi.org/10.1016/j.compag.2020.105384.
Zhang, Y., Tian, Y., Zheng, C., Zhao, D., Gao, P., Duan, K., 2019. Segmentation of apple point clouds based on ROI in RGB images. Inmateh - Agric. Eng. 59, 209–218. https://doi.org/10.35633/INMATEH-59-23.
Zhang, Z., Flores, P., Igathinathane, C., Naik, D.L., Kiran, R., Ransom, J.K., 2020b. Wheat lodging detection from UAS imagery using machine learning algorithms. Remote Sens. 12, 1838. https://doi.org/10.3390/rs12111838.
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 59, 1165–1180. https://doi.org/10.13031/trans.59.11737.
Zhang, Z., Igathinathane, C., Li, J., Cen, H., Lu, Y., Flores, P., 2020c. Technology progress in mechanical harvest of fresh market apples. Comput. Electron. Agric. 175, 105606. https://doi.org/10.1016/j.compag.2020.105606.
Zhang, Z., Pothula, A.K., Lu, R., 2018b. A review of bin filling technologies for apple harvest and postharvest handling. Appl. Eng. Agric. 34, 687–703. https://doi.org/10.13031/aea.12827.
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.