Received: 29 July 2023 | Revised: 14 November 2023 | Accepted: 8 January 2024

DOI: 10.1002/rob.22297

RESEARCH ARTICLE

WILEY

Performance evaluation of newly released cameras for fruit detection and localization in complex kiwifruit orchard environments

Xiaojuan Liu[1] | Xudong Jing[1] | Hanhui Jiang[1] | Shoaib Younas[1] | Ruiyang Wei[1] | Haojie Dang[1] | Zhenchao Wu[1] | Longsheng Fu[1,2,3]

[1]College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi, China

[2]Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi, China

[3]Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling, Shaanxi, China

Correspondence

Longsheng Fu, College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China.
Email: fulsh@nwafu.edu.cn and longsheng.fu@outlook.com

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 32371999, 32171897; National Foreign Expert Project, Ministry of Science and Technology, China, Grant/Award Numbers: DL2022172003L, QN2022172006L; Key Research and Development Program of Shaanxi, China, Grant/Award Number: 2023JBGS-21

Abstract

Consumer RGB-D and binocular stereo cameras were applied to fruit detection and localization. However, few studies are documented on performance comparison of newly released cameras under same scene in complex orchard. This study evaluates performance of consumer RGB-D and binocular stereo cameras based on YOLOv5x for kiwifruit detection and localization and selection of optimal one with better application in complex orchard environment. Firstly, Azure Kinect, RealSense D435, and ZED 2i cameras were employed to capture images of kiwifruit canopies. Subsequently, YOLOv5x was applied to train and detect kiwifruits and calyxes in the images. Meanwhile, an overlap-partitioning detection strategy was applied on kiwifruit and calyx detection. Additionally, spatial coordinate transformation was performed by integrating camera's extrinsic parameters and depth map generated by each camera. Finally, three-dimensional coordinates of calyxes were calculated and compared with ground truth, followed by localization accuracy of calyxes were analyzed. Results show that YOLOv5x obtained mean average precision of 93.2%, 91.3%, and 95.8% for three cameras on kiwifruit and calyx detection, respectively. Overlap-partitioning detection strategy improved the calyx detection and significantly increased average precision by 13.00%, 16.30%, and 7.70%, respectively. The mean absolute deviation of calyx coordinates on Y-axis is relatively high for ZED 2i at 8.44 mm in comparison of 6.67 mm for Azure Kinect, while RealSense D435 achieved minimum of 10.42 mm on X-axis and 18.33 mm on Z-axis. Average spatial localization speed of calyxes in one image was 0.164 s, 0.037 s, and 0.062 s for Azure Kinect, RealSense D435, and ZED 2i, respectively. These results indicate the excellent performance of RealSense D435 than Azure Kinect and ZED 2i in kiwifruit orchard, which could be a valuable reference for other orchards to select a camera with high precision localization capacity.

KEYWORDS

binocular stereo camera, consumer RGB-D camera, overlap-partitioning detection, performance evaluation, YOLOv5x

1 | INTRODUCTION

Machine vision systems have been extensively used in sustainable production process of modern agriculture, which enables detection and localization of target fruits, such as automatic fruit harvesting, orchard yield prediction, orchard yield mapping, and fruit growth monitoring (Fu, Feng, et al., 2020; Sun et al., 2022). Detection and localization of target fruits are certainly important components during the process of automatic fruit harvesting. However, accurate detection and localization of target fruits is a highly challenging task in complex orchard environment, since target fruits are easily affected by various of illumination conditions or low-resolutions, covered by foliage or branches, and overlapping of adjacent fruits (Fu, Gao, et al., 2020; Li et al., 2020; Lin et al., 2019). Therefore, machine vision systems are critical for accurate detection and localization of fruits.

Depth cameras, as key equipment in the machine vision system from fruit harvesting robots, provide reliable visual information for fruit detection and localization. According to camera's depth measuring principle, these depth cameras are categorized into consumer RGB-D and binocular stereo cameras. Represented by Kinect and RealSense series of cameras, consumer RGB-D cameras utilize Time-of-Flight (ToF) or Active Infrared Stereoscopy (AIRS) technique to acquire depth information (Giancola et al., 2018; Tadic et al., 2022), while binocular stereo cameras adopt passive measurement based on stereoscopic technology (Sosa-León & Schwering, 2022). There was increasing trend found in popularity of depth cameras in the field of machine vision system due to their ability to capture depth information for determining the object size, location, and distance between object and camera (Vit & Shani, 2018). However, different types of depth cameras have their own advantages and disadvantages. It is crucial to choose a camera with better application capacity in complex orchard environment for fruit detection and localization.

Currently, several studies used consumer RGB-D and binocular stereo cameras to detect and locate single fruit in specific field environment. Some researchers applied Kinect v2 in detection and localization of litchi, citrus and tomato (Li et al., 2020; Lin et al., 2019; Li et al., 2023). Similarly, others adopted RealSense R200 and D435 to detect and locate strawberry and apple (Wang et al., 2022; Xiong et al., 2019). Moreover, a few researchers used binocular vision cameras (ZED mini, ZED 2 and ZED v2) in detection and localization of tomato, oil-seed camellia fruit and grape (Hsieh et al., 2021; Tang et al., 2023; Xiao et al., 2023). Above-mentioned studies mainly were conducted in structured environments. However, there are few studies on performance comparison of consumer RGB-D and binocular stereo cameras for fruit detection and localization under same scene in unstructured (complex orchard) environments. Therefore, it is necessary to compare newly released cameras for fruit detection and localization in complex orchard environment.

Kinect series of cameras have gained fame and rapid development was observed for different applications in machine vision system, including fruit detection and localization. Li et al. (2020) adopted Kinect v2 to detect and locate fruit-bearing branches of multiple litchi clusters simultaneously, and obtained detection accuracy of 83.33% and localization accuracy of 17.29° ± 24.57°. Lin et al. (2019) used Kinect v2 and developed a reliable algorithm to detect and locate citrus, which achieved F1 score of 0.92 and localization errors in x, y, z directions were 7.0 ± 2.5 mm, −4.0 ± 3.0 mm, and 13.0 ± 3.0 mm, respectively. Similarly, Li, He, et al. (2021) applied Kinect v2 in the same direction on tea shoots in a developed algorithm, which obtained picking success rate of 83.18% for tea shoots and average localization time of about 24 ms for each tea shoot. Furthermore, Nguyen et al. (2016) estimated three-dimensional (3D) position and diameter of red or bicolored apples using Kinect v1 and obtained localization estimation error below 10 mm in all coordinate axes.

Similarly, Intel's RealSense series of cameras have shown immense potential in diverse agricultural applications with characteristics of small size, low cost, easy durability and high accuracy. Wang et al. (2022) adopted two RealSense D435 to accurately recognize and retrieve apples in field environment, which obtained harvesting success rate ranging from 70% to 85%. Li, Sun, et al. (2021) proposed a fast detection and localization method of Longan fruits using an UAV equipped with RealSense D455 in mountain orchard, which achieved localization time of 0.5 s per image for Longan fruit strings and fruit branches. Zhang et al. (2021) developed a deep learning-based apple detection and localization system using RealSense D435i, which picked 80 apples with efficiency of 82.47%. Xiong et al. (2019) used RealSense R200 combining with depth information and intrinsic matrix in the calculation of 3D location of strawberry, and showed average localization error of 13.3 mm. Ge et al. (2020) proposed a 3D shape completion method equipped with RealSense D435 for strawberry localization and obtained 0.61 and 5.7 mm intersection over union and center deviation, respectively.

Additionally, binocular stereo cameras offer potential advantages mainly in fruit localization applications, and possess the advantage of less affected by illumination and light reflection because of resemblance with human eyes. In consequences of these beneficial properties, Tang et al. (2023) applied ZED 2 on oil-seed camellia fruit and found localization errors of 23.57 ± 7.42 mm and 23.52 ± 7.43 mm under sunlight and shading conditions, respectively. In another study of ZED mini on beef tomato in greenhouse average error was 0.5 cm (Hsieh et al., 2021). However, in the application of ZED 2 in citrus orchard observed average error was 12.3 mm (Chen et al., 2021). A visual localization system with two charge-coupled devices was used for litchi clusters localization, which obtained maximum and minimum depth error of 5.08 and 1.96 cm (Xiong et al., 2018).

Since, only few studies have been documented to compare and evaluate different cameras in structured environments. Neupane et al. (2021) compared eight depth cameras for localization and sizing in fruit tree orchard. The findings of this study recommended that Azure Kinect or Blaze 101 are operational in sunlight and orchard conditions. Vit and Shani (2018) evaluated four low cost RGB-D cameras in outdoor agricultural phenotyping tasks and found that RealSense D435 provided a viable tool for close range phenotyping tasks in field. Whereas, given that technology limitations of each depth camera, Condotta et al. (2020) conducted a research in agricultural applications of five depth cameras containing three

available technologies (Structured light, ToF, Stereoscopy), which indicated ToF and stereoscopy technology were recommended for indoor and outdoor applications, respectively. Scientists put a lot of efforts to provide evaluation performance evidence while high-performance evaluation of newly released cameras for fruit detection and localization under same scene in complex orchard environment need to explore. As such, there is an imperative need for comparing and evaluating consumer RGB-D and binocular stereo cameras for fruit detection and localization in complex orchard environment.

Kiwifruit has been widely cultivated in China and becoming popular in consumers. As one of the semi-tropical fruits, it possesses significant amount of nutrients and vitamin C. Kiwifruit is commercially grown on sturdy support structures such as T-bar trellis and pergolas, where the T-bar trellis cultivation pattern is widespread in China (Fu, Gao, et al., 2020). Kiwifruit orchards are extremely complex, and in addition to the fruit there are other components such as leaves, branches, wires and stems. Fruits are often occluded by leaves, adjacent fruits, and wires in canopy images. In addition, there are variable illumination conditions in orchards. It is worth noting that, calyx of each kiwifruit is visible and independent unlike kiwifruit in canopy images (Song et al., 2021). Hence, it is a desirable strategy to identify midpoint of calyx as localization point for end-effectors in fruit localization under complex orchard environment.

This paper aimed to evaluate the performance of consumer RGB-D and binocular stereo cameras for fruit detection and localization. Additionally, it discussed the selection of a camera with better application capacity in complex orchard environment and performance evaluation of newly released cameras (i.e., Azure Kinect, RealSense D435 and ZED 2i) based on You Only Look Once version 5 extra large (YOLOv5x) for fruit detection and localization in kiwifruit orchard. The rest of this paper further split as follows: Section 2 describes materials and methods including three newly released cameras, 3D coordinates measurement, dataset acquisition and processing, fruit detection and localization. Section 3 presents the obtained results of performance evaluation of three cameras according to fruit and calyx detection, calyx localization, and discusses relevant issues. Finally, the conclusions of this work are summarized in Section 4.

2 | MATERIALS AND METHODS

Our research evaluated the performance of three newly released cameras for localization accuracy of calyxes and observed their application capacity under same scene in kiwifruit orchard. The overall schematic workflow of this study is shown in Figure 1.

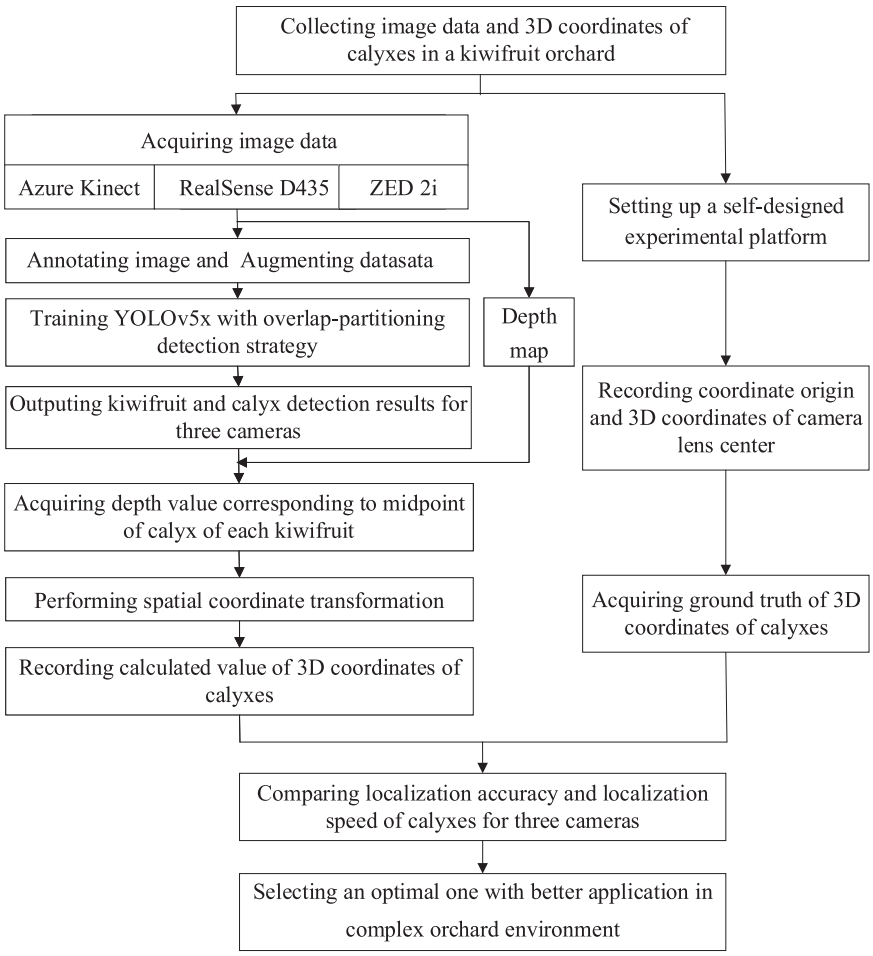

FIGURE 1 Overall schematic workflow.

2.1 | RGB-D and binocular stereo cameras

In addition to the parameters provided on the official website, manually measured parameters and principles of three cameras is required for fruit localization. In this study, all three cameras were fixed on a same bracket in an experimental setup, and their relative location is shown in Figure 2.

Main parameters of three cameras are listed in Table 1. Focal distance and principal point are two intrinsic parameters of cameras for calculating 3D coordinates of calyxes. According to our knowledge, each camera enables to take RGB images and corresponding depth maps. When a depth map is aligned with its corresponding RGB image, depth values of the aligned depth map can be read. However, it is worthy to notice that, not every pixel coordinate has depth values in the aligned depth map, such as on shiny, bright, transparent, and distant regions (Zhang & Funkhouser, 2018). Thus, a searching strategy for depth values was performed within the neighborhood of pixel coordinates, and searching range was set to the maximum limited of 15 pixels based on kiwifruit pixel size. Notably, depth measuring principles of three cameras are distinguishing, which highly related with fruit localization. Details on depth measuring principles of three cameras are introduced as follows.

2.1.1 | Azure Kinect

Azure Kinect is commercialized to provide data for a wide area of applications in agriculture (McGlade et al., 2020; Suo et al., 2021). The camera adopted ToF technology and calculated depth information of target object from the time it takes for near-infrared rays to travel from emission to reception. ToF measuring principle is shown in Figure 3.

Given the frequency of infrared emitter is f, the wavelength of NIR light is λ, the speed of light is c, the time from emission to reception is t, the phase difference is $\Delta\varphi$, and the distance (d) can be calculated by Equations (1) and (2), respectively.

$$T = \frac{\lambda}{c} = \frac{1}{f}, \tag{1}$$

$$d = \frac{1}{2}c \times t = \frac{1}{2}\left(n\lambda + \frac{\Delta\varphi}{2\pi}\lambda\right). \tag{2}$$

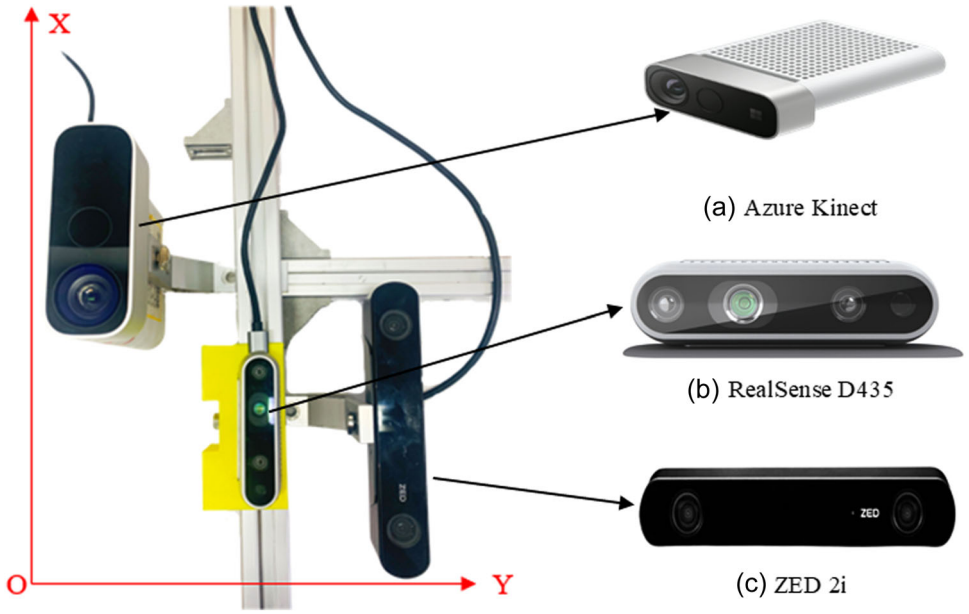

FIGURE 2 Relative location of each camera. (a) Azure Kinect, (b) RealSense D435, and (c) ZED 2i.

TABLE 1 Main parameters of three cameras.

Camera	Range (m)	Resolution (RGB and depth) (pixels)	Focal distance (pixel)		Principal point (pixel)	
			f_x	f_y	c_x	c_y
Azure Kinect	0.5 ~ 3.9	1920 × 1080	972.9558	977.6413	980.2478	515.1505
RealSense D435	0.28 ~ 3.0	1280 × 720	645.7929	651.5319	649.1980	368.6035
ZED 2i	0.3 ~ 20.0	1280 × 720	525.9000	526.6500	554.1000	311.9500

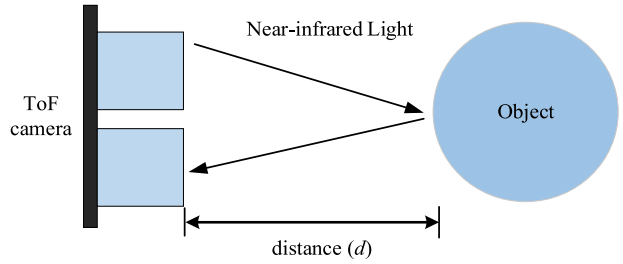

FIGURE 3 Time-of-Flight measuring principle.

It should be noted that an aligned depth map is a set of coordinate values (Z_c), which are integrated with each pixel (u, v) in its corresponding RGB image to generate 3D coordinate P_c (X_c, Y_c, Z_c). Besides, after acquiring Z_c from the aligned depth map, X_c and Y_c were calculated by Equations (4) and (5), respectively. Where, (u, v) refers to pixel coordinate, f_x, f_y, c_x, c_y represents the focal distance and the principal point of three cameras, respectively. However, the depth map generated by low-cost Kinect devices has some problems, such as depth missing areas on the edge of target object, RGB images' edge do not aligned with the edge of target object, and more depth noise in depth map (Wang et al., 2016).

$$Zc = d, \tag{3}$$

$$X_c = \frac{(u - c_x)}{f_x} \times Z_c, \tag{4}$$

$$Y_c = \frac{(v - c_y)}{f_y} \times Z_c. \tag{5}$$

2.1.2 | RealSense D435

RealSense D435 provides the wider field of view among Intel RealSense sensors. Moreover, it enables to yield acceptable depth measurement results while target objects are at a distance of few meters (Tadic et al., 2022). AIRS technology was adopted by RealSense D435 to acquire depth information with the addition of a pair of stereo infrared sensors (stereo IR camera), an infrared laser emitter (IR projector), and RGB camera. It should be mentioned that depth value indicating object depth (Z) is determined based on parallel plane of RealSense D435 doing capturing, not in relation to actual range (R) of the object from RealSense D435 (Tadic et al., 2022). In depth measurement method, a diagram of depth (Z) measurement versus range (R) is shown in Figure 4.

Similarly, as one of the RGB-D cameras, a depth map generated by RealSense D345 also has depth missing to some extent. Besides, one more thing to emphasize here is that calculation method of 3D coordinates of calyxes for RealSense D435 is the same as Equations (3) to (5).

2.1.3 | ZED 2i

ZED 2i is designed for outdoor applications and challenging agricultural environments (ZED Product Portfolio, 2022), which should be applied for further research in robotic vision field (Tadic et al., 2022). As a depth camera, it collects depth information based on stereoscopic technology (Sosa-León & Schwering, 2022; Tadic et al., 2022). It is important to note that depth measuring principle of ZED 2i is distinguished from Azure Kinect and RealSense D435. That is, a binocular camera is employed to capture 3D data in a scene and measure disparity of an object using a stereo matching algorithm (Zhang et al., 2021), though calculate depth map according to parameters of the depth camera (Tadic et al., 2022; ZED Product Portfolio, 2022). Depth measuring principle using a binocular stereo camera is displayed in Figure 5. Where, P_c (X_c, Y_c, Z_c) is a spatial point in the scene, parallel to RGB sensors in left and right lenses of the depth camera. P_l (X_l, Y_l) and P_r (X_r, Y_r) are the projection of spatial point P_c on left and right lens. O_l and O_r indicate origins of the two lens coordinate systems, respectively.

Where, P_c (X_c, Y_c, Z_c) is a spatial point in the scene, parallel to RGB sensors in left and right lenses of the depth camera. P_l (X_l, Y_l) and P_r (X_r, Y_r) are the projection of spatial point P_c on left and right lens. O_l and O_r indicate origins of the two lens coordinate systems, respectively. According to principle of similar triangles in plane geometry, the related coordinates of P_l (X_l, Y_l) and P_r (X_r, Y_r) are defined as shown in Equations (6) to (8), respectively, and Y_l is equal to Y_r. The camera coordinate P_c (X_c, Y_c, Z_c) can be obtained by Equations (9) to (11), respectively. Overall, it's a crucial step to accurately locate target fruits to find coordinates of corresponding

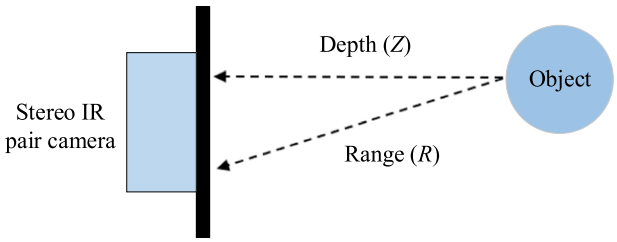

FIGURE 4 Diagram of depth (Z) measurement versus range (R) (Tadic, 2019).

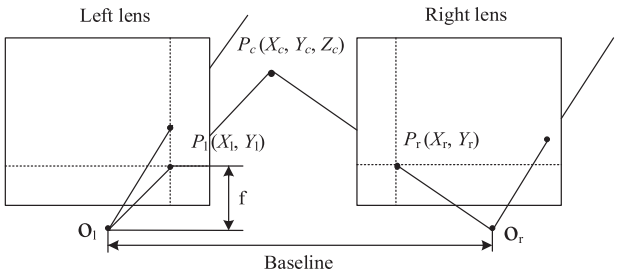

FIGURE 5 Depth measuring principle using a binocular stereo camera. Black intersection of a line between P_c and origin O_l and imaging plane of the left lens is P_l, while P_r is the intersection of a line between P_c and origin O_r and imaging plane of the right lens.

points in left and right images according to Equations (6) to (11). Where, baseline refers to the disparity between left and right images and f represents focal distance length of a binocular camera.

$$X_l = \frac{X_c}{Z_c} \times f, \tag{6}$$

$$X_r = \frac{X_c - \text{Baseline}}{Z_c} \times f, \tag{7}$$

$$Y_l = \frac{Y_c}{Z_c} \times f, \tag{8}$$

$$X_c = \frac{\text{Baseline}}{X_r - X_l} \times X_l, \tag{9}$$

$$Y_c = \frac{\text{Baseline}}{X_r - X_l} \times Y_l, \tag{10}$$

$$Z_c = \frac{\text{Baseline}}{X_r - X_l} \times f. \tag{11}$$

Owing to the baseline shown in Figure 5, a visual field area appears in left and right images, where the pixels located in this area cannot get feature matching on the other image, resulting in uncalculated disparity. So, depth missing also appears in a depth map generated by ZED 2i (Song et al., 2021).

2.2 | 3D coordinates measurement

Aforementioned performance of three cameras was evaluated using localization accuracy of calyxes. In this work, midpoints of calyxes of each kiwifruit in a canopy image was determined as localization points for end-effectors picking during fruit localization task.

Ground truth of 3D coordinates of calyxes was measured by a laser rangefinder (VCHON H-40, JinYun, China) on a self-designed experimental platform, as shown in Figure 6. The experimental platform was set up approximately 1 m below dense-foliage kiwifruit canopy. Initially, five level gauges and four micro height adjusting devices were applied to adjust the platform on horizontal position. Then, a coordinate paper was pasted on the platform, and coordinate origin was recorded. Later on, 10 kiwifruits within field of view of each camera were randomly marked. A laser rangefinder was employed to find the midpoint of calyx in each marked kiwifruit, thus Z-coordinate of the midpoint of calyx was recorded. In the third step, 3D coordinates of calyxes were recorded in combination with the coordinate paper. Finally, all the above steps were repeated by changing different areas. In total, 3D coordinates of the calyxes of 89 kiwifruits were recorded in 10 different areas.

2.3 | Dataset acquisition and processing

2.3.1 | Image acquisition

Kiwifruit images were captured by three cameras in an orchard at two different times in morning and afternoon under different illumination conditions. For this study, three cameras fixed to the experimental platform separately were connected with a computer and applied to collect images from the bottom of dense-foliage kiwifruit canopy in a vertical upward manner, as shown in Figure 6. Kiwifruit images, that is, original RGB images and depth maps were collected under same scene in late October 2021 and 2022 during harvesting seasons of common "Hayward" cultivar at Meixian Kiwifruit Experimental Station (34°07′39′ N, 107°59′50′ E, 648 m

FIGURE 6 A self-designed experimental platform of this work. Where, three cameras are fixed on the same bracket to ensure that relative location of three cameras remains unchanged and cameras' lens are horizontal. (a) Laser rangefinder, (b) level gauge, and (c) micro-adjusting holder.

in altitude), Northwest A&F University, Shaanxi, China. The captured images were saved in "JPG" format with corresponding resolution.

2.3.2 | Image dataset and annotation

A total of 3496 images were collected using three cameras, out of which 3456 images were used for training network and 40 images were used for calyx localization. The original images (160 RGB-D images and 40 pairs of binocular images) were randomly divided into training set and test set with 4:1 ratio. Then, seventeen dataset augmentation methods were adopted to improve overall model performance, including brightness transformation, contrast transformation, chroma transformation, sharpness transformation, blurring transformation in gaussian and smooth, image mirroring in horizontal and vertical axis, image rotation in 90°, 180°, and 270°, where the thresholds for brightness transformation, contrast transformation, and chroma transformation were set to 0.8, 1.2 and 1.5, respectively. After dataset augmentation, the training set was augmented from 64 images to 1152 images for each camera.

Acquired images were manually annotated with LabelImg1.8 (https://github.com/tzutalin/labelImg). Kiwifruit and calyx in acquired images were annotated as rectangles with labels "Kiwi" and "Calyx," respectively. Figure 7 shows some labeling examples of kiwifruit and calyx in an image. After annotating one image, annotation files were saved in corresponding "xml" format. The original and augmented image datasets are available on https://github.com/fu3lab/Kiwifruit_performance-evaluation_newly-released-cameras-_images.

2.4 | Fruit and calyx detection

2.4.1 | Network and training hyperparameters

You Only Look Once version 5 extra large (YOLOv5x) shows great potential for object detection in machine vision. Compared with the other three versions of YOLOv5 series, YOLOv5x has large complexity (model depth and layer channel), relative fast detection speed and high detection accuracy (Cao et al., 2023; Xu et al., 2022). Accordingly, considering the advantages of this model, YOLOv5x was employed for kiwifruit and calyx detection in this paper.

Experiments were performed based on PyTorch framework with version 1.10.2 based on a desktop computer equipped with AMD Ryzen 7 5800×8-Core Processor (3.80 GHz) CPU, Nvidia GeForce GTX 3080 Ti 12 G GPU (10240 CUDA cores), and 64 GB of memory, running on a Windows 10 64-bit system. Software tools included CUDA 10.1, CUDNN 11.0, Python 3.9, and OpenCV 4.1. YOLOv5x was applied for training kiwifruit and calyx detection under PyTorch framework. The network input size was 640 × 640 pixels, with a batch size of 8. Stochastic gradient descent was applied for training with a momentum of 0.937 and a weight decay of 0.0005. A value of 0.001 was set as initial learning rate of the network. Iterations were set to 350 to analyze training process. During training YOLOv5x, as a machine learning method, transfer learning referred to a pretrained model being reused in another task and led to faster and more accurate training results (Fu, Majeed, et al., 2020; Li et al., 2022; Suo et al., 2021).

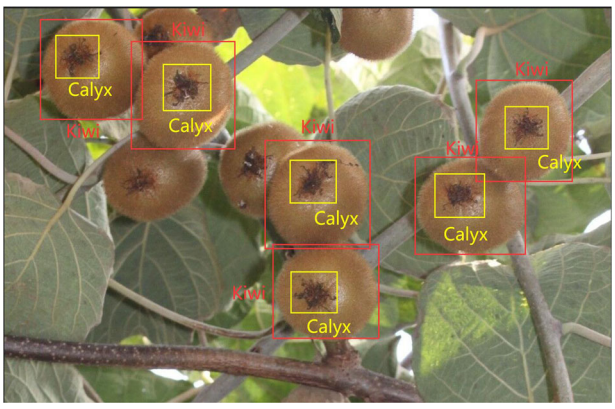

FIGURE 7 Labeling examples of kiwifruit and calyx. Kiwifruit is labeled as "Kiwi" using a red rectangle, and calyx is labeled as "Calyx" using a yellow rectangle.

2.4.2 | Overlap-partitioning detection strategy

Accurate calyx detection of each kiwifruit is essential to ensure subsequent calyx localization. Since the calyx accounts for a small percentage of pixels in kiwifruit images, traditional detection methods are less effective for detection of small target objects. An overlap-partitioning detection strategy was adopted for improving the calyx detection in this paper. The process of overlap-partitioning detection strategy is shown in Figure 8.

Initially, the size of a large image was read, and a certain overlap ratio of two adjacent small images was set to 20%. Meanwhile, the large image was partitioned to small images of 640 × 640 pixels from top-left to bottom-right according to the input size of YOLOv5x network. Afterward, over-partitioned space of small images in last row or last column was panned and padded. Finally, each partitioned small image was trained and detected separately, and the detection results of partitioned small images were stitched into the large image. For example, a large image captured by Azure Kinect was partitioned into small images as shown Figure 9.

2.5 | Calyx localization

2.5.1 | Camera calibration

Acquiring camera's intrinsic parameters and calyx position in RGB images is two key steps of calyx localization. In the first step, accurate camera calibration is crucial for obtaining intrinsic parameters, as presented in Table 1. Once intrinsic parameters were obtained, the calyx position in camera coordinate system can be accurately

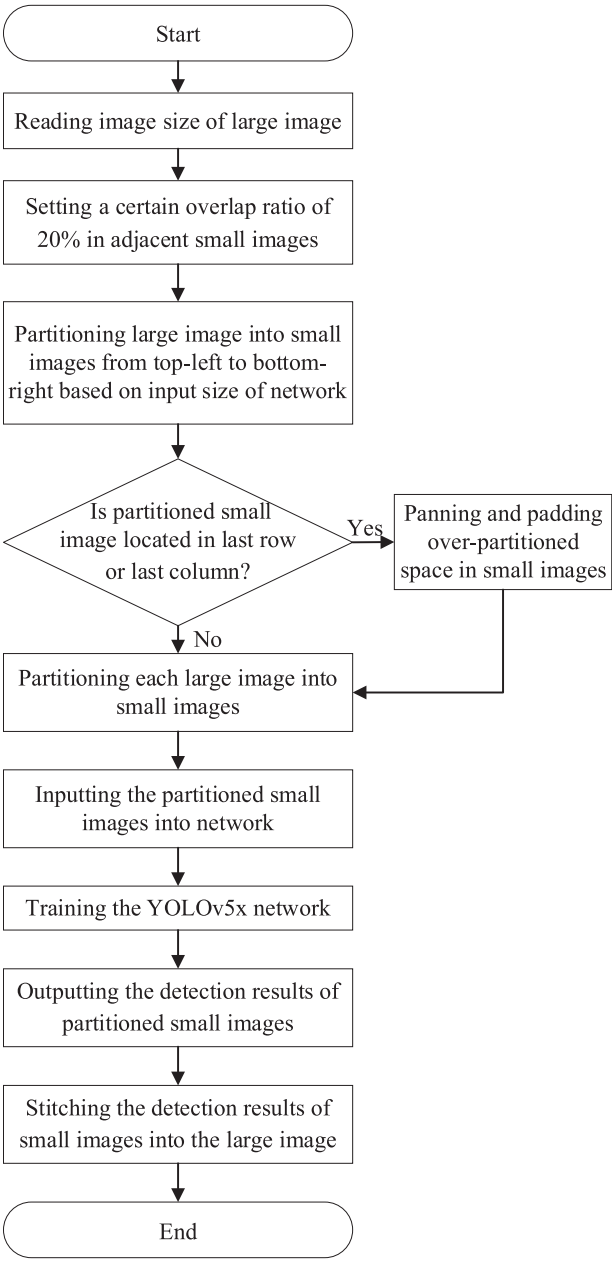

FIGURE 8 The process of overlap-partitioning detection strategy.

calculated in 3D space (Liu, 2020). In this paper, three cameras were calibrated by adopting a calibration method proposed by Zhang (1998). Each camera captured 20 calibration board images for calibration from different angles and positions to ensure robustness and accuracy of calibration results.

2.5.2 | Spatial coordinate transformation

After camera calibration, spatial coordinate transformation from camera system to world coordinate system is another essential step to acquire 3D coordinates of calyxes. The motivation and necessity of spatial coordinate transformation include two aspects. First, working space of kiwifruit harvesting robot is in world coordinate system, but coordinates of midpoint of calyx obtained by each camera are in camera coordinate system. Second, to facilitate comparison of localization accuracy for three cameras, it is necessary to unify coordinate system and compare the midpoint of calyx under same coordinate origin.

The relative location of each camera was demonstrated and a uniform coordinate origin was determined, which were utilized for comparing localization accuracy of calyxes. In the next step, extrinsic parameters of three cameras were calculated based on camera position relative to coordinate origin, including rotation matrix (R_c) and translation matrix (T_c), and the subscript (C) in R_c and T_c refers to each camera (Liu, 2020). Furthermore, coordinate transformation from camera to world was accomplished by combining the previously mentioned camera coordinate (shown in 2.1.1 to 2.1.3 for details) with cameras extrinsic parameters. The coordinate P_w (X_w, Y_w, Z_w) of calyx in world coordinate were obtained, as shown in Equation (12) and transformation process from camera to world coordinate in Figure 10.

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = RC \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + TC, \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} RC & TC \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}, RC : 3 \times 3, TC : 3 \times 1. \tag{12}$$

2.6 | Performance evaluation

The performance of kiwifruit and calyx detection was evaluated by average precision (*AP*), mean average precision (*mAP*), and average detection speed. All samples were divided into four groups: *TP* (true positive), *FN* (false negative), *FP* (false positive), and *TN* (true negative) according to the combinations of true and predicted class (Gao et al., 2022). First of all, *AP* is calculated by precision (*P*) and recall (*R*) of the detection network. *P* is a measured value of detection results relevancy, while *R* is a measured value of how many truly relevant detection results are returned. *P* and *R* are defined as shown in Equation (13) and Equation (14), respectively.

$$P = \frac{TP}{TP + FP}, \tag{13}$$

$$R = \frac{TP}{TP + FN}. \tag{14}$$

The AP_i refers to area under P_i and R_i curve of the i^{th} class (i.e., kiwifruit or calyx in this paper), which is a standard for measuring sensitivity of the network to target object, and an indicator of global performance reflection of network (Gao et al., 2020), shown in Equation (15). Next, *mAP* is defined in Equation (16) as the *AP* of the i^{th} class. The higher *AP* and *mAP*, the better detection results of convolutional neural network for a given object (Gao et al., 2020;

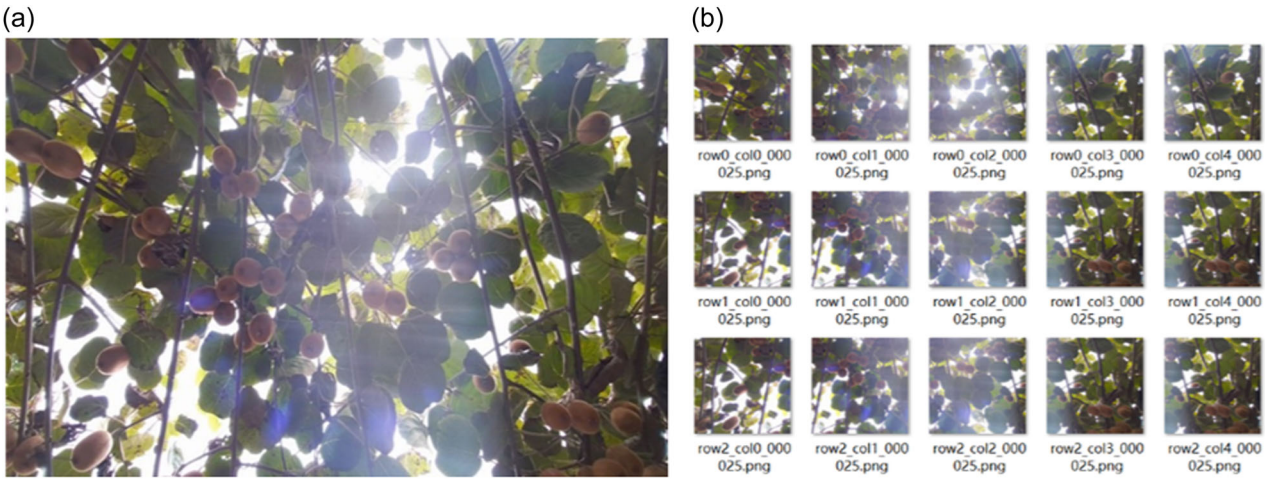

FIGURE 9 Example of a large image was partitioned into small images. (a) A large image captured by Azure Kinect, and (b) small images partitioned from a large image after adopting overlap-partitioning detection strategy.

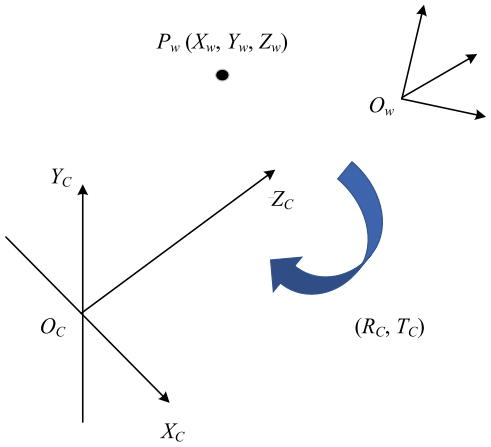

FIGURE 10 Transformation from camera coordinate to world coordinate. Where, O_c-$X_cY_cZ_c$ is the camera coordinate; O_w-$X_wY_wZ_w$ is the world coordinate; P_w (X_w, Y_w, Z_w) is calyx coordinate in world coordinate; R_c and T_c refer to rotation matrix and translation matrix, respectively.

Zhang et al., 2020). Besides, average detection speed was applied to evaluate performance of YOLOv5x.

$$AP_i = \int_0^1 P_i(R_i)dR_i, \tag{15}$$

$$mAP = \frac{1}{2}\sum_{i=1}^{2} AP_i, \tag{16}$$

where *TP* represents the number of correctly detected kiwifruits or calyxes, *FP* indicates the number of incorrectly detected kiwifruits or calyxes, and *FN* refers to the number of missed kiwifruits or calyxes.

The performance of calyx localization was assessed by localization accuracy and localization speed. These evaluation indicators and the analysis of corresponding results can further reflect performance comparison of RGB-D and binocular stereo cameras. In terms of localization accuracy, it refers to mean absolute deviation in one coordinate-axis (MAD_{c_axis}), where the one coordinate-axis represents X-axis or Y-axis or Z-axis of the world coordinate, and the subscript (C) in MAD_{c_axis} represents each camera. It is worth noting that MAD_{c_axis} was also considered as a crucial performance evaluation indicator for three cameras in this paper, and the definition of MAD_{c_axis} is shown in the following Equation (17):

$$MAD_{c\_axis} = \frac{|CV_{c\_axis} - GT|}{N}, \tag{17}$$

where CV_{c_axis} and GT are calculated values in one coordinate-axis and the ground truth of 3D coordinates of calyxes, respectively. N represents the total number of kiwifruits in calyx localization images.

3 | RESULTS AND DISCUSSION

3.1 | Fruit and calyx detection performance

3.1.1 | Before adopting overlap-partitioning detection strategy

Performance evaluation metrics, including *AP*, *mAP*, and detection speed, are mainly employed for verifying YOLOv5x model. Table 2 shows kiwifruit and calyx detection results using YOLOv5x before adopting overlap-partitioning detection strategy for three cameras. The application of Azure Kinect, RealSense D435, and ZED 2i in kiwifruit and calyx detection obtained *mAP*s of 93.2%, 91.3%, and 95.8% respectively. Moreover, *AP* of 99.7% was high kiwifruit detection in three cameras. YOLOv5x has excellent performance on kiwifruit detection in trained images, in contrast to its relatively low performance on calyx detection. And small difference was observed in terms of average detection speed.

The *mAP*s and *AP*s determined that ZED 2i obtained significantly high detection results followed by Azure Kinect in trained images,

while RealSense D435 had relatively low detection results. The fact is that low resolution of Azure Kinect caused relative low detection performance. Resolution of captured images by Azure Kinect was 1920 × 1080 pixels, while the resolution of RealSense D435 and ZED 2i image was 1280 × 720 pixels. Initially, original images were compressed into 640 × 640 pixels during detection process, and images were more compressed as size increases. Moreover, less discriminations were found after image compression because of small pixel areas in one image. Therefore, the detection results for Azure Kinect are low profiled in comparison of ZED 2i. Whereas, in case of RealSense D435, the reason of quiet low detection is that a RGB image was compressed as the depth map aligned with the RGB image during model training (Vijayanagar et al., 2014). In addition, it is also possible that the RGB image was scaled or changed after alignment of depth map with corresponding RGB image, resulting in poor image quality. Also, coordinates information without depth values in the aligned depth map was mapped to the corresponding RGB image, or even the RGB information was not obtained, as shown in Figure 11.

3.1.2 | After adopting overlap-partitioning detection strategy

The combination of YOLOv5x model with overlap-partitioning detection strategy improved the calyx detection. Detection time is directly proportional to image size and increases with original image size because there are a large number of small images. Table 3 shows kiwifruit and calyx detection results using YOLOv5x after adopting overlap-partitioning detection strategy for three cameras. The *mAP*s and *AP*s were remarkably improved compared with the detection results in Table 2, especially the *AP*s on calyx detection. The *mAP* was increased by 6.50%, 8.20%, and 3.90% on kiwifruit and calyx detection, and *AP* was increased by 13.00%, 16.30%, and 7.70% on calyx detection for three cameras, respectively.

The average detection speed of YOLOv5x for Azure Kinect was slower than the other cameras shown in Table 3, which might be due to high number of small images (15 small images) in Azure Kinect than RealSense D435 (6 small images) and ZED 2i (6 small images) after adopting overlap-partitioning detection strategy. To the best of our knowledge, fruit-harvesting robots adopted a two-step strategy. The first is to detect fruits in field view of a camera, and the second is a harvesting sequence and path planning for robot arms and end-effector to harvest fruits (Gao et al., 2020). It has been reported that kiwifruit harvesting robot takes approximately 2.16 s to pick a kiwifruit with one-arm gripper (Williams et al., 2020). Hence, it is acceptable to spend 896.33 ms for a kiwifruit harvesting robot to detect a canopy image in variable and complex orchard environment.

Kiwifruit and calyx detection results of all the cameras revealed that the detection method missed the detection and collected false identification information of calyx before adopting overlap-partitioning detection strategy, while this problem was resolved by adopting the strategy. Detection results of both stages are shown in Figure 12. Where, Figure 12a,c,e represent detection results before adopting overlap-partitioning detection strategy, and the rest of subfigures represent results after adopting the strategy. To further visualize comparison results for both stages, in Figure 12a,c,e, red dotted ellipse examples indicate that calyx was incorrectly identified as kiwifruit, and yellow dotted ellipse examples indicate that the calyx was not identified, but only kiwifruit was identified. The correct detection results are shown in Figure 12b,d,f, namely the examples of red or yellow solid ellipse. It is obviously found that kiwifruit and calyx detection in three cameras using YOLOv5x after adopting the overlap-partitioning detection strategy is greatly improved.

Lastly but most significantly, it should be mentioned that the calyx detection was significantly enhanced, even though the strategy

TABLE 2 Kiwifruit and calyx detection results using YOLOv5x before adopting overlap-partitioning detection strategy for three cameras.

Camera	*mAP* (%)	*AP* (%)		Average detection speed (ms/image)
		Kiwifruit	Calyx	
Azure Kinect	93.2	99.7	86.7	59.69
RealSense D435	91.3	99.8	82.9	56.25
ZED 2i	95.8	99.7	92.0	59.88

FIGURE 11 Example of an aligned RGB image after alignment of depth map with corresponding RGB image captured by RealSense D435.

TABLE 3 Kiwifruit and calyx detection results using YOLOv5x after adopting overlap-partitioning detection strategy for three cameras.

Camera	*mAP* (%)	*AP* (%)		Average detection speed (ms/image)
		Kiwifruit	Calyx	
Azure Kinect	99.7	99.7	99.7	896.33
RealSense D435	99.5	99.7	99.2	551.60
ZED 2i	99.7	99.7	99.7	615.20

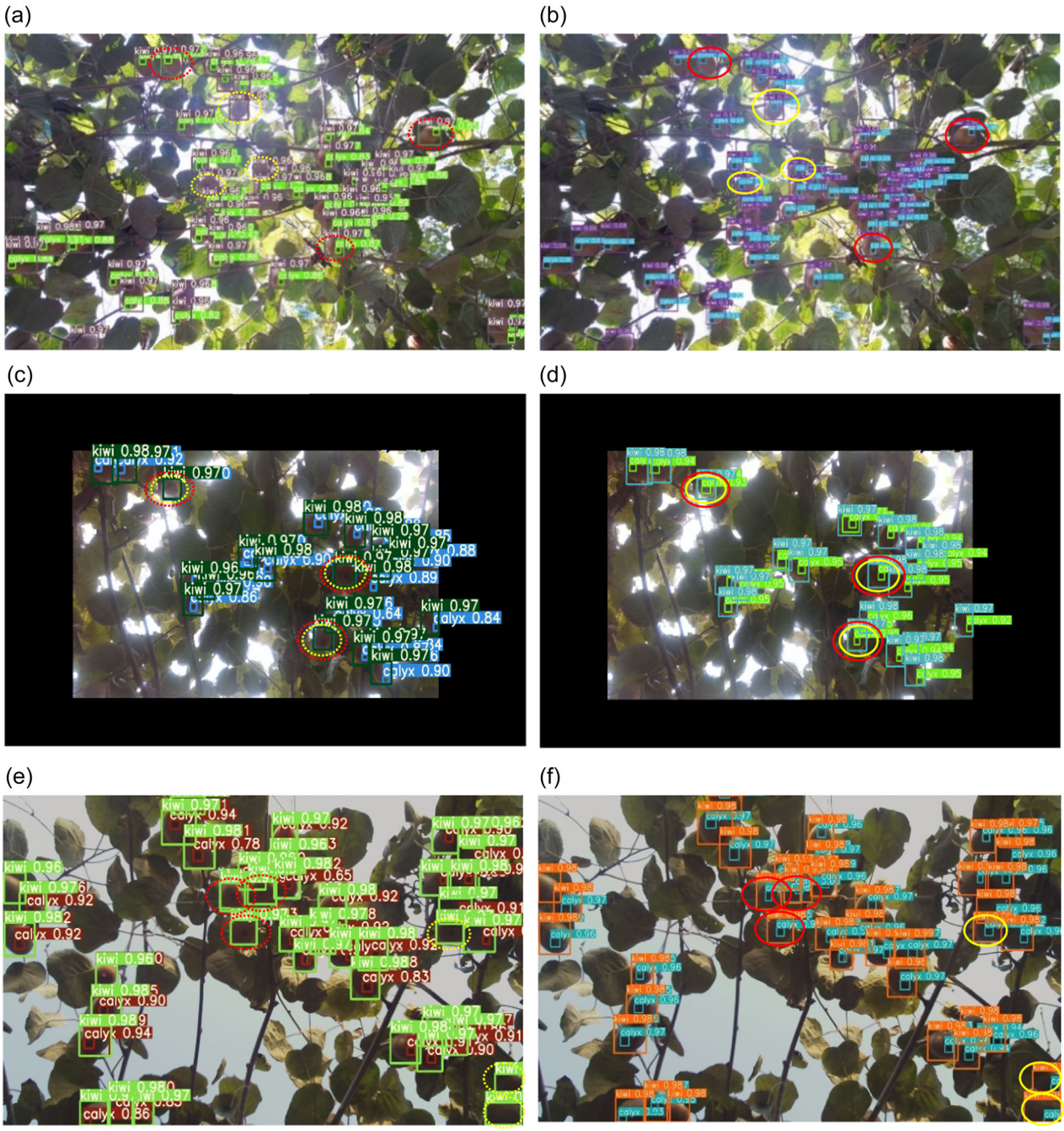

FIGURE 12 Detection results before and after adopting overlap-partitioning detection strategy for three cameras. (a), (c), and (e) represent detection results before adopting overlap-partitioning detection strategy, and the rest of subfigures represent the detection results after adopting the strategy for Azure Kinect, RealSense D435, and ZED 2i, respectively. Red dotted ellipse examples indicate that calyx was incorrectly identified as kiwifruit, and yellow dotted ellipse examples indicate that the calyx was not identified, but only kiwifruit was identified. Red or yellow solid ellipse examples indicate that the correct detection results.

decreased detection speed as previously mentioned, as can be seen the examples of yellow solid ellipse in Figure 12b,d,f. It can be concluded that the overlap-partitioning detection strategy solved image compression problem due to large image size and small input size limited by hardware devices. Overall, the above-mentioned calyx detection and speed fulfill real-time detection requirement of kiwifruit and the subsequent calyx localization.

3.2 | Calyx localization performance

3.2.1 | Localization accuracy

The localization accuracy of calyxes for the three cameras were analyzed, as shown in Figure 13. Azure Kinect obtained the highest MAD_{c_axis} on *X*-axis, and may be because the resolution of acquired

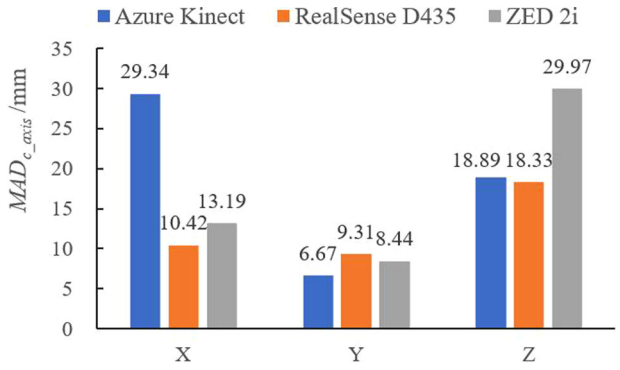

FIGURE 13 MAD_{c_axis} of calyxes for the three cameras.

images was inconsistent with other cameras. ZED 2i obtained the maximum MAD_{c_axis} of 29.97 mm on Z-axis and a relatively high MAD_{c_axis} of 13.19 mm on *X*-axis, while there was no clear difference of MAD_{c_axis} on Y-axis in relation of other cameras. However, RealSense D435 reached the lowest MAD_{c_axis} on X and Z axes and the relatively high MAD_{c_axis} of 9.31 mm on Y-axis, and expressed the less difference from Azure Kinect. Results showed that the MAD_{c_axis} for RGB-D cameras basically smaller than a binocular camera, and RealSense D435 achieved preferable localization accuracy of calyxes on three axes.

Certainly, there are some possible reasons on deviation of MAD_{c_axis} of calyxes for three cameras. The first reason was difference in depth measuring principle of three cameras. It's reported that Azure Kinect was observed to bring greater deviation (McGlade et al., 2020), depth deviation of RealSense D435 is less than that of Kinect series when the camera distance from the object was one meter (Ahn et al., 2019; Liu, 2020; Vit & Shani, 2018). As can be also seen from Figure 13, localization deviation of Azure Kinect is larger than that of RealSense D435 on *Z*-axis. Secondly, inconsistent image resolution may result in the larger MAD_{c_axis} on *X* and *Z* axes for Azure Kinect and ZED 2i. Finally, a certain deviation was found between midpoint of detection rectangular and calyx, which can cause different deviation of MAD_{c_axis}.

Although the localization accuracy of calyxes is not optimal compared with previous studies (Chen et al., 2021; Lin et al., 2019; Nguyen et al., 2016; Xiong et al., 2019), the MAD_{c_axis} of calyxes for three cameras is acceptable and satisfies design requirement of 30 mm deviation for kiwifruit harvesting robot arm (Liu, 2020). Comprehensive analysis demonstrated that RealSense D435 obtained the best performance in calyx localization and highlighted the excellent application in complex orchard environment.

3.2.2 | Localization speed

Overlap-partitioning detection strategy is the premise of acquiring 3D coordinates of calyxes. In this study, the localization speed of calyxes consists of detection speed and spatial localization speed. In localization speed analysis, test set (16 images) were used to calculate detection speed of calyxes, and the dataset from different areas (10 images) were applied in calculation of spatial localization speed of calyxes. An important point is that results of average detection speed were obtained after adopting overlap-partitioning detection strategy. Average spatial localization speed consists of two parts, one is speed of indexing depth values in corresponding depth map, while the second part includes speed of locating calyxes. Table 4 lists the calyx localization speed results using YOLOv5x.

TABLE 4 Calyx localization speed results using YOLOv5x for three cameras (mean ± standard deviation).

Camera	Resolution (pixels)	Average detection speed of calyxes (s/image)	Average spatial localization speed of calyxes (s/image)
Azure Kinect	1920 × 1080	0.896 ± 0.002[a]	0.164 ± 8.231E−04[d]
RealSense D435	1280 × 720	0.552 ± 0.003[b]	0.037 ± 3.530E−06[e]
ZED 2i	1280 × 720	0.615 ± 0.007[c]	0.062 ± 5.760E−05[f]

Note: Same letters in the third and fourth columns represent no significant difference at the 0.05 level for three cameras.

Significant differences were found in calyx localization speed among three cameras, demonstrated in Table 4. The average detection speed might be connected with the number of small images input by YOLOv5x after adopting overlap-partitioning detection strategy. Whereas, for average spatial localization speed of calyxes, this is mainly due to the consistent neighborhood search from the process of indexing the depth values in corresponding depth map. As previously mentioned, depth map generated by RGB-D or binocular stereo cameras suffered from a lot of missing values. Therefore, the neighborhood searching increased the average localization time of calyxes for two RGB-D cameras compared with relevant studies (Li, He, et al., 2021; Li, Sun, et al., 2021). Nevertheless, Kinect v2 was employed by Lin et al. (2019) to detect and locate a citrus fruit in an average of 1.25 s, which took more localization time than that of calyxes in a kiwifruit image for Azure Kinect. Overall, RealSense D435 performed well in terms of localization speed of calyxes as compared to the rest of cameras.

4 | CONCLUSIONS

This study remarkably evaluated the performance and selection of newly released cameras (i.e., Azure Kinect, RealSense D435, and ZED 2i) based on YOLOv5x for kiwifruit detection and localization for the better application capacity. Detection results (*mAP*, *AP*, and detection speed) achieved from three cameras were satisfied. YOLOv5x had accurate and simultaneous detection efficiency for kiwifruits and calyxes. After adopting overlap-partitioning detection strategy, the *mAP*s were increased by 6.50%, 8.20%, and 3.90% on kiwifruit and calyx detection, and *AP*s were increased by 13.00%, 16.30%, and

7.70% on calyx detection for Azure Kinect, RealSense D435, and ZED 2i, respectively. This increasing trend greatly improved the detection of small target objects. Although average detection speed of YOLOv5x was slow after adopting overlap-partitioning detection strategy, it satisfies the requirement of real-time detection of kiwifruit for harvesting robot. Moreover, Azure Kinect obtained the minimum MAD_{c_axis} of 6.67 mm on Y-axis, RealSense D435 achieved the minimum MAD_{c_axis} of 10.42 mm X-axis and 18.33 mm on Z-axis, and ZED 2i obtained the relatively low MAD_{c_axis} of 8.44 mm on Y-axis, which all satisfies design requirement of 30 mm tolerance for kiwifruit harvesting robot arm. The average spatial localization speed of the calyxes indicates RealSense D435 has the better application in complex orchard environments. In future, this study could provide a reference information for the selection of high-performance camera in other orchards.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32371999, 32171897); Key Research and Development Program of Shaanxi, China (2023JBGS-21); National Foreign Expert Project, Ministry of Science and Technology, China (DL2022172003L, QN2022172006L).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available https://github.com/fu3lab/Kiwifruit_performance-evaluation_newly-released-cameras-_images.

ORCID

Longsheng Fu http://orcid.org/0000-0003-3253-2637

REFERENCES

Ahn, M., Chae, H., Noh, D., Nam, H. & Hong, D. (2019) Analysis and noise modeling of the Intel RealSense D435 for mobile robots. In: *2019 16th International Conference on Ubiquitous Robots (UR)*. Jeju, Korea (South): IEEE, pp. 707–711. Available from: https://doi.org/10.1109/URAI.2019.8768489

Cao, Y., Chen, J. & Zhang, Z. (2023) A sheep dynamic counting scheme based on the fusion between an improved-sparrow-search YOLOv5x-ECA model and few-shot deepsort algorithm. *Computers and Electronics in Agriculture*, 206, 107696. Available from: https://doi.org/10.1016/j.compag.2023.107696

Chen, M., Tang, Y., Zou, X., Huang, Z., Zhou, H. & Chen, S. (2021) 3D global mapping of large-scale unstructured orchard integrating eye-in-hand stereo vision and SLAM. *Computers and Electronics in Agriculture*, 187, 106237. Available from: https://doi.org/10.1016/j.compag.2021.106237

Condotta, I.C.F.S., Brown-Brandl, T.M., Pitla, S.K., Stinn, J.P. & Silva-Miranda, K.O. (2020) Evaluationof low-cost depth cameras for agricultural applications. *Computers and Electronics in Agriculture*, 173, 105394. Available from: https://doi.org/10.1016/j.compag.2020.105394

Fu, L., Feng, Y., Wu, J., Liu, Z., Gao, F., Majeed, Y. et al. (2020) Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precision Agriculture*, 22, 754–776. Available from: https://doi.org/10.1007/s11119-020-09754-y

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. *Computers and Electronics in Agriculture*, 177, 105687. Available from: https://doi.org/10.1016/j.compag.2020.105687

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. *Biosystems Engineering*, 197, 245–256. Available from: https://doi.org/10.1016/j.biosystemseng.2020.07.007

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G. et al. (2022) A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. *Computers and Electronics in Agriculture*, 197, 107000. Available from: https://doi.org/10.1016/j.compag.2022.107000

Gao, F., Fu, L., Zhang, X., Majeed, Y., Li, R., Karkee, M. et al. (2020) Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. *Computers and Electronics in Agriculture*, 176, 105634. Available from: https://doi.org/10.1016/j.compag.2020.105634

Ge, Y., Xiong, Y. & From, P.J. (2020) Symmetry-based 3D shape completion for fruit localisation for harvesting robots. *Biosystems Engineering*, 197, 188–202. Available from: https://doi.org/10.1016/j.biosystemseng.2020.07.003

Giancola, S., Valenti, M. & Sala, R. (2018) *A survey on 3D cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies*. SpringerBriefs in computer science. Cham, Switzerland: SpringerBerlin/Heidelberg, pp. 1–96. Available from: https://doi.org/10.1007/978-3-319-91761-0

Hsieh, K.W., Huang, B.Y., Hsiao, K.Z., Tuan, Y.H., Shih, F.P., Hsieh, L.C. et al. (2021) Fruit maturity and location identification of beef tomato using R-CNN and binocular imaging technology. *Journal of Food Measurement and Characterization*, 15, 5170–5180. Available from: https://doi.org/10.1007/s11694-021-01074-7

Li, D., Sun, X., Elkhouchlaa, H., Jia, Y., Yao, Z., Lin, P. et al. (2021) Fast detection and location of longan fruits using UAV images. *Computers and Electronics in Agriculture*, 190, 106465. Available from: https://doi.org/10.1016/j.compag.2021.106465

Li, G., Fu, L., Gao, C., Fang, W., Zhao, G., Shi, F. et al. (2022) Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and Euclidean distance. *Computers and Electronics in Agriculture*, 201, 107342. Available from: https://doi.org/10.1016/j.compag.2022.107342

Li, J., Tang, Y., Zou, X., Lin, G. & Wang, H. (2020) Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots. *IEEE Access*, 8, 117746–117758. Available from: https://doi.org/10.1109/ACCESS.2020.3005386

Li, Q., Sun, X., Jiang, H., Wu, A., Fu, L. & Li, R. (2023) Design and test of intelligent spraying unmanned vehicle for greenhouse tomato based on YOLOv4-tiny. *Journal of Intelligent Agricultural Mechanization*, 4(2), 44–52. Available from: https://doi.org/10.12398/j.issn.2096-7217.2023.02.005

Li, Y., He, L., Jia, J., Lv, J., Chen, J., Qiao, X. et al. (2021) In-field tea shoot detection and 3D localization using an RGB-D camera. *Computers and Electronics in Agriculture*, 185, 106149. Available from: https://doi.org/10.1016/j.compag.2021.106149

Lin, G., Tang, Y., Zou, X., Li, J. & Xiong, J. (2019) In-field citrus detection and localisation based on RGB-D image analysis. *Biosystems Engineering*, 186, 34–44. Available from: https://doi.org/10.1016/j.biosystemseng.2019.06.019

Liu, Z. (2020) *Kiwifruit detection and localization methods based on multi-source information fusion*. [Master Thesis, Northwest A&F University]. Shaanxi, China. Available from: https://doi.org/10.27409/d.cnki.gxbnu.2020.000944

McGlade, J., Wallace, L., Hally, B., White, A., Reinke, K. & Jones, S. (2020) An early exploration of the use of the Microsoft Azure Kinect for estimation of urban tree diameter at breast height. *Remote Sensing Letters*, 11, 963–972. Available from: https://doi.org/10.1080/2150704X.2020.1802528

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, 1780. Available from: https://doi.org/10.3390/agronomy11091780.

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. *Biosystems Engineering*, 146, 33–44. Available from: https://doi.org/10.1016/j.biosystemseng.2016.01.007

Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R. et al. (2021) Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. *Computers and Electronics in Agriculture*, 181, 105933. Available from: https://doi.org/10.1016/j.compag.2020.105933

Sosa-León, V.A.L. & Schwering, A. (2022) Evaluating automatic body orientation detection for indoor location from skeleton tracking data to detect socially occupied spaces using the Kinect v2, Azure Kinect and Zed 2i. *Sensors*, 22, 3798. Available from: https://doi.org/10.3390/s22103798

Sun, M., Xu, L., Luo, R., Lu, Y. & Jia, W. (2022) Fast location and recognition of green apple based on RGB-D image. *Frontiers in Plant Science*, 13, 864458. Available from: https://doi.org/10.3389/fpls.2022.864458

Suo, R., Gao, F., Zhou, Z., Fu, L., Song, Z., Dhupia, J. et al. (2021) Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking. *Computers and Electronics in Agriculture*, 182, 106052. Available from: https://doi.org/10.1016/j.compag.2021.106052

Tadic, V., 2019. *Intel RealSense D400 series product family datasheet*. Satan Clara, CA: New Technologies Group, Intel Corporation. Document number: 337029-013.

Tadic, V., Toth, A., Vizvari, Z., Klincsik, M., Sari, Z., Sarcevic, P. et al. (2022) Perspectives of RealSense and ZED depth sensors for robotic vision applications. *Machines*, 10, 183. Available from: https://doi.org/10.3390/machines10030183

Tang, Y., Zhou, H., Wang, H. & Zhang, Y. (2023) Fruit detection and positioning technology for a *Camellia oleifera* C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Systems with Applications*, 211, 118573. Available from: https://doi.org/10.1016/j.eswa.2022.118573

Vijayanagar, K.R., Loghman, M. & Kim, J. (2014) Real-time refinement of kinect depth maps using multi-resolution anisotropic diffusion. *Mobile Networks and Applications*, 19, 414–425. Available from: https://doi.org/10.1007/s11036-013-0458-7

Vit, A. & Shani, G. (2018) Comparing RGB-D sensors for close range outdoor agricultural phenotyping. *Sensors*, 18, 4413. Available from: https://doi.org/10.3390/s18124413

Wang, X., Kang, H., Zhou, H., Au, W. & Chen, C. (2022) Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. *Computers and Electronics in Agriculture*, 193, 106716. Available from: https://doi.org/10.1016/j.compag.2022.106716

Wang, Z., Song, X., Wang, S., Xiao, J., Zhong, R. & Hu, R. (2016) Filling Kinect depth holes via position-guided matrix completion. *Neurocomputing*, 215, 48–52. Available from: https://doi.org/10.1016/j.neucom.2015.05.146

Williams, H., Ting, C., Nejati, M., Jones, M.H., Penhall, N., Lim, J. et al. (2020) Improvements to and large-scale evaluation of a robotic kiwifruit harvester. *Journal of Field Robotics*, 37, 187–201. Available from: https://doi.org/10.1002/rob.21890

Xiao, Z., Luo, L., Chen, M., Wang, J., Lu, Q. & Luo, S. (2023) Detection of grapes in orchard environment based on improved YOLO-v4. *Journal of Intelligent Agricultural Mechanization*, 4(2), 35–43. Available from: https://doi.org/10.12398/j.issn.2096-7217.2023.02.004

Xiong, J., He, Z., Lin, R., Liu, Z., Bu, R., Yang, Z. et al. (2018) Visual positioning technology of picking robots for dynamic litchi clusters with disturbance. *Computers and Electronics in Agriculture*, 151, 226–237. Available from: https://doi.org/10.1016/j.compag.2018.06.007

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. *Computers and Electronics in Agriculture*, 157, 392–402. Available from: https://doi.org/10.1016/j.compag.2019.01.009

Xu, X., Wang, L., Shu, M., Liang, X., Ghafoor, A.Z., Liu, Y. et al. (2022) Detection and counting of maize leaves based on two-stage deep learning with UAV-based RGB image. *Remote Sensing*, 14(21), 5388. Available from: https://doi.org/10.3390/rs14215388

ZED Product Portfolio. (2022) *Stereolabs product portfolio and specifications*. Revision 1. Orsay: Stereolabs.

Zhang, J., Zhang, Y., Wang, C., Yu, H. & Qin, C. (2021) Binocular stereo matching algorithm based on MST cost aggregation. *Mathematical Biosciences and Engineering*, 18, 3215–3226. Available from: https://doi.org/10.3934/mbe.2021160

Zhang, Y. & Funkhouser, T. (2018) Deep depth completion of a single RGB-D Image. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake, UT: IEEE, pp. 175–185. https://doi.org/10.1109/CVPR.2018.00026

Zhang, Z. (1998) A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334. Available from: https://doi.org/10.1109/34.888718

Zhang, Z., Flores, P., Igathinathane, C., L. Naik, D., Kiran, R. & Ransom, J.K. (2020) Wheat lodging detection from UAS imagery using machine learning algorithms. *Remote Sensing*, 12, 1838. Available from: https://doi.org/10.3390/rs12111838

How to cite this article: Liu, X., Jing, X., Jiang, H., Younas, S., Wei, R., Dang, H. et al. (2024) Performance evaluation of newly released cameras for fruit detection and localization in complex kiwifruit orchard environments. *Journal of Field Robotics*, 41, 881–894. https://doi.org/10.1002/rob.22297