3D Vision-based Perception and Length Estimation of Green Asparagus for Selective Harvesting

Jiajun Xu

Michigan State University

Yuzhen Lu

luyuzhen@msu.edu

Michigan State University https://orcid.org/0000-0002-3636-1137

Research Article

Keywords: Artificial intelligence, Green asparagus, Machine vision, Point cloud processing, Selective harvesting

Posted Date: January 3rd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5746943/v1

Additional Declarations: The authors declare no competing interests.

3D Vision-based Perception and Length Estimation of Green Asparagus for Selective Harvesting

Jiajun Xu, Yuzhen Lu*

Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

Abstract

Green asparagus is a labor-intensive vegetable crop for harvesting. Rising labor costs and seasonal labor shortages are threatening the sustainability and profitability of the asparagus industry in the United States (U.S.), making harvesting automation more urgently needed than ever. A multitude of challenges, however, exist in developing practically variable automated harvesting technology for green asparagus. This study describes a novel effort aimed at the development and evaluation of a 3D vision-based asparagus perception system for selective harvesting of U.S. green asparagus. Three different types of 3D cameras were evaluated indoors for asparagus detection and length estimation. A time-of-flight technology-based camera yielded the best accuracy and was deployed on a mobile vision system for imaging asparagus in diverse field conditions. A new annotated dataset alongside ground-truth length measurements for 1008 spears was created for asparagus perception algorithm development and evaluation. Among three small-scale YOLO (v8, v9, and v10) models, YOLOv8s achieved the best F1 score of 0.849. The YOLOv8s-based pipeline point cloud processing algorithms including segmentation, clustering, and outlier removal, achieved a percentage accuracy of 89.4% (with the corresponding error of about 2.3 cm) in spear length estimation, and the averaged base point localization error of 1.4 cm. The entire algorithm pipeline for spear detection and localization could be run at about 3.3 frames per second on an onboard computer. The harvest eligibility analysis of detected spears showed a detection rate of 95.8% of harvestable spears when only the length criterion (greater than 8 inches or 20.3 cm) was applied and a detection rate of 84.5% when both the length

*Corresponding author

Email address: luyuzhen@msu.edu (Yuzhen Lu)

and base point localization (error less than 2 cm) criteria were considered. The 3D vision-based perception system in this study is promising for reliable and real-time asparagus perception in production fields.

Keywords: Artificial intelligence, Green asparagus, Machine vision, Point cloud processing, Selective harvesting

---

1. Introduction

Asparagus is one of the important perennial vegetable crops grown worldwide, with global production value exceeding $13.4 billion (FAOSTAT, 2024). In the U.S., the production of asparagus, predominately green asparagus, requires frequent harvests (between 50 and 70) (Cembali et al., 2005), due to the rapid growth of spears (the editable portions aboveground), with some varieties growing at a rate of up to 50 mm in length per day (Arndt et al., 1997). Green asparagus is harvested manually and selectively for a season of typically 6-8 weeks. In 2023, the U.S. harvested 15,700 acres of green asparagus, generating a farmgate value exceeding $72 million, with over 85% of the crop value due to the fresh market (USDA-NASS, 2024). Michigan is the leading state in green asparagus production in the U.S., representing 55% of the total acreage planted nationwide. Because of frequent, manual harvesting and rising labor costs, harvesting labor for green asparagus is the most costly production activity, accounting for over 55% of the total production costs (Bardenhagen et al., 2023). The high labor cost alongside the stiff competition with foreign producers (Herrygers, 2024) is threatening the sustainability of the U.S. asparagus industry, with the nationwide planted area of asparagus declining by more than 80% since 2000. Therefore, for producers to remain in production and competitive, there is an urgent need to develop automated harvesting technology for green asparagus.

Research into automated harvesting of green asparagus has a long history, with the most intensive efforts made to mechanize asparagus harvesting during the 1950s to 1960s (Kepner, 1959; Knicely, 1962; Mears and Moore, 1969; Moore and Carpenter, 1968). Overall, two types of approaches, selective and non-selective, were researched in developing mechanical asparagus harvesters (Chen et al., 2010; MARNE et al., 2021). Non-selective harvesters cut all spears indiscriminately, regardless of their maturity (length), each time the field is traversed, while selective harvester aims to replicate the manual harvesting process by harvesting the spears taller than a preselected height and leaving the shorter ones for future harvests, improving marketable yield and quality and minimizing non-saleable produce (Kepner, 1971). Selective harvesters require dedicated perception systems based on either non-vision or vision-based sensors to detect harvest-eligible spears, followed by actuating properly designed mechanisms to perform harvesting operations. For successful harvesting, accurate asparagus localization and length estimation are key factors that influence the performance of selective harvesters. Additionally, harvest aids have been introduced to improve ergonomic comfort for laborers, but still rely on humans to perform harvesting.

Different sensing technologies have been applied for asparagus perception in selective harvesting, including non-vision and vision-based methods. Arndt et al. (1997) reported on an automated selective asparagus harvesting robot developed in Austria, which was equipped with fiber-optic photoelectric sensors mounted at a preset height from the ground for spear detection. However, field tests revealed that the harvester could not adjust to the types of beds used in the U.S. and failed to cut stalks shorter than 20 cm since it would not break the sensor beam. Existing commercial prototypes such as the Kim Haws harvester (Haws, 1997, 2018) and the Geiger Lund harvester (Lund, 1985, 2022) relies on optical detectors with pulsed lasers, mounted at a preselected height above the ground, to sense spears of the corresponding height as the harvester moves down the crop row. However, such sensing systems require no sophisticated data processing but have low localization accuracy and may be unreliable in variable field light conditions. Sakai et al. (2013) used laser scanners for asparagus perception in the greenhouse environment, which took about 2 seconds per spear, but the performance of the sensors was limited by insufficient scanning resolution, leading to mediocre accuracy of spear detection. Similar studies were reported on using laser scanners for localizing the spears of the desired height (Irie et al., 2009; Irie and Taguchi, 2014). Recently, Šlajpah et al. (2023) reported on the use of a 2D (two-dimensional) LiDAR (light detection and ranging) sensor on a mobile system for asparagus perception, where a point cloud of the spears was generated by combining laser scanner data with the platform odometry data. The LiDAR-based system achieved an 88% detection rate for harvest-ready spears (with height of 15 cm or greater) at a platform speed of 0.06 m/s in field testing, with the harvesting success rate down to 77% due to clustered growth and localization issues.

Compared with non-vision-based sensing, imaging technology conceivably offers enhanced perception capabilities. In early studies, monochromatic cameras were employed to capture 2D images of asparagus spears, which were binarized using an empirically determined threshold to separate spears from the soil background (Baylou et al., 1984; Grattoni et al., 1994). Localization of spears was achieved using a single camera coupled with a perspective transform (Humburg, 1991; Humburg and Reid, 1991) and stereo cameras (Baylou et al., 1984; Grattoni et al., 1994; Grattoni and Guiducci, 1990). However, spear segmentation by thresholding could be easily affected by unstructured field environments and varying lighting conditions, making precise spear localization challenging. Three-dimensional (3D) vision has advanced rapidly over the last 20 years and become increasingly popular in agricultural applications (Miranda et al., 2023; Vázquez-Arellano et al., 2016), with the prevalence of consumer-grade RGB-D (red-green-blue-depth) cameras (Drouin and Seoud, 2020). Different 3D imaging techniques exist, including time-of-flight (ToF), stereovision, and structured light imaging, among which TOF imaging sensors appear to be promising for unstructured field environments (Peebles et al., 2018), given their relative insensitivity to ambient light conditions (Vázquez-Arellano et al., 2016; Zanuttigh et al., 2016).

Leu et al. (2017) employed a ToF RGB-D camera (Microsoft Kinect v2) for asparagus perception in a selective harvest robot, where the vision module recognized and tracked eligible asparagus (longer than 15 cm) in generated 3D point clouds. Field tests of the robot at a traveling speed of 0.2 m/s gave a success rate of about 90%, but limitations existed in detecting the spears that were close together (e.g., spears separated by 10 cm or less in the lateral direction) and harvesting them without collateral damage. Peebles et al. (2020) combined an RGB camera and a ToF camera (Microsoft Kinect v2) for the detection and localization of asparagus spears, respectively. Their robotic system achieved 92.3% accuracy in cutting targeted, obtainable spears at a platform speed of 0.33 m/s, but it showed poor performance when tested in low-light conditions. It is noted that these earlier ToF sensors have relatively low depth map resolution, restricting their ability to localize asparagus spears in close proximity. Mu et al. (2024) reported on the use of an active stereo vision camera (Intel RealSenseTM D435i) for asparagus perception, which achieved decent accuracy in greenhouse conditions. Despite these studies, there has been no dedicated research on the performance evaluation of different types of RGB-D cameras for informed sensor selection and optimized asparagus perception in field conditions. Most previous studies did not report the performance metrics in spear length estimation, rendering it difficult to objectively assess the effectiveness of asparagus perception.

For robust asparagus perception, apart from deploying proper imaging sensors, equally important is the development of computer vision algorithms for the detection, localization, and length estimation of asparagus. Traditional image processing methods (e.g., thresholding, morphological operations) (Humburg, 1991; Humburg and Reid, 1991; Rigney et al., 1992), tend to have limited generalization performance in natural field conditions. Deep learning techniques, especially convolutional neural networks (CNNs), have recently gained unprecedented popularity for handling visual recognition tasks in harvesting automation of specialty crops (Suresh Kumar and Mohan, 2023; Vasconez et al., 2020).

Peebles et al. (2019) evaluated Faster R-CNN and SSD (single shot multibox detector) for asparagus detection and achieved an F1 score of up to 0.73. In a follow-up study, the authors reported on mixed detection accuracy by Faster R-CNN, ranging from 57.7% to 90.9% in different field-testing conditions (Peebles et al., 2024). Zhang et al. (2022) proposed an asparagus detection method for white asparagus based on YOLOv5 optimized for embedded devices, achieving an F1 score of 0.975. Different modifications made to the YOLOv5 were reported for improved asparagus detection with mAP (mean average precision) exceeding 95% (Hong et al., 2023; Zhang et al., 2024). CCN-based instance segmentation models were also exploited for asparagus detection (Liu et al., 2022; Mu et al., 2024). For instance, Mu et al. (2024) made architectural modifications to YOLOv8n for subsoil cutting point localization of asparagus, achieving an F1 score of 98.2%, a cutting-point localization accuracy of over 99.0%, and a relative length estimation error of about 3%. However, the authors only sampled a small set of asparagus spears for performance assessment in length estimation. It is important to note that these studies (Zhang et al., 2022, 2024; Liu et al., 2022; Hong et al., 2023; Mu et al., 2024), some of which focused on white asparagus, were conducted for asparagus grown in protected, relatively structured facilities (e.g., greenhouses) instead of open fields where the green asparagus is generally grown in the U.S.

With a long-term goal to develop vision-guided selective harvesting technology for the U.S. asparagus, this study aimed to evaluate different 3-D vision techniques for asparagus perception and develop computer vision algorithms with state-of-the-art deep learning for asparagus detection and length estimation in open field conditions. To the best of our knowledge, this study represents the very first effort to benchmark different RGB-D cameras for asparagus length estimation in simulated settings and thereafter conduct extensive imagery with the selected 3-D sensor for asparagus grown in natural, open field conditions in the U.S., alongside ground-truth length measurements for over 1000 individual spears. The open field conditions, especially where asparagus spears are densely populated with cluttered field backgrounds, elevate the level of challenges or difficulty in asparagus detection compared to that in greenhouse-based studies reported in the literature. Although previous research examined CNNs for asparagus detection, rapid advancements in real-time object detectors, especially YOLOs, warrant a new instigation of their performance in asparagus detection using the imagery acquired in field conditions.

Specific objectives of this study were therefore four-fold, i.e., to 1) evaluate and compare the performance of three different types of 3D imaging sensors in the detection and length measurements of green asparagus in simulated, indoor environments, 2) create a large dataset containing labeled 3D imagery and ground-truth length measurements of spears for the U.S. green asparagus acquired a mobile machine vision system with a selected 3D imager in diverse production fields, 3) develop and evaluate a computer vision algorithm pipeline with pound cloud analysis, powered by state-of-the-art YOLO models, for asparagus spear detection, length estimation, and base point localization with 3D field imagery, and 4) further conduct harvest eligibility analysis for detected asparagus spears.

2. Materials & Methods

2.1. Mobile Vision Platform Setup

A lightweight, ground-based mobile platform, which was originally built for weed imaging in our prior study (Xu et al., 2024), was deployed for image acquisition of green asparagus. As schematically shown in Figure 1, the platform consists of a wheeled aluminum frame enclosed with opaque black panels that minimize the effects of variable ambient lighting and other environmental conditions (e.g., high temperatures), a configuration that will be adapted in the future harvesting robot design to ensure consistent imaging quality. With the enclosure, an RGB-D camera pointing downward is mounted on a rotational bracket, allowing angle adjustments to make the camera face forward as needed, and the bracket itself is mounted on a vertically adjustable linear track, enabling flexible adjustment of the height of the camera from the ground. Surrounding the camera, two 12V LED strips are attached to both sides of the frame to provide consistent, uniform lighting throughout the interior during imaging. The platform is equipped with an onboard computer (Jetson AGX Orin$^{\mathrm{TM}}$, NVIDIA, Santa Clara, CA, USA), along with supporting peripherals such as a monitor and a keyboard. The computer is compact yet powerful (64 GB memory with 2,048 CUDA cores and 64 Tensor cores) for machine vision and artificial intelligence applications. The wheels of the platform operate with four independent gear motors, which can be controlled by a remote-control device. A 12 V battery

(50 Ah) is used for powering the motorized platform as well as the computer and monitor via a DC/AC inverter, ensuring efficient and reliable operation in various field conditions.

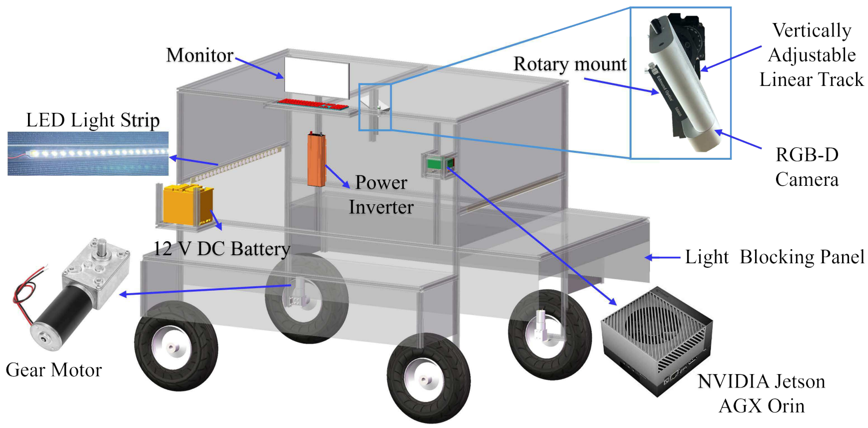

Figure 1: A custom-built mobile machine vision platform with an RGB-D (red-green-blue-depth) camera for asparagus imaging.

2.2. Indoor imaging

An indoor imaging experiment was first conducted to evaluate three different types of RGB-D cameras and determine the one with the best sensing performance in length estimation of asparagus, which would be utilized for subsequent field imaging and performance evaluation. As shown in Figure 2, a simulated asparagus bed with dimensions 103 cm (length) $\times$ 56 cm (width) is prepared in a wood frame filled with sand to mimic the typical sandy soil growing environment of asparagus, where asparagus spears are planted upright with varied heights. Here, a total of 140 asparagus spears, purchased from local grocery stores, were used for the imaging experiment. Individual spears were manually marked at the insertion position and pre-measured for their length before being arranged in batches of 15–25 spears for each image acquisition session.

For the indoor camera performance evaluation, as shown in Figure 3, the Z-height of each asparagus in the soil bin was manually measured using a ruler with a resolution of 1 mm to serve as the ground truth of its height. The Z-height was determined by marking the insertion point of the spear into the soil and measuring the vertical length above the insertion point. Such length definition was commonly used in non-vision-based asparagus perception systems where light-beam sensors are positioned at a

Figure 2: Simulated asparagus bed for indoor experiments.

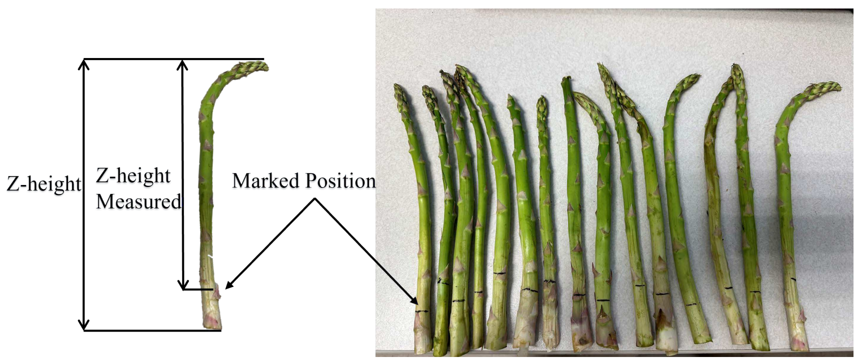

Figure 3: The Z-height and marked position of each asparagus.

predetermined height to detect spears for selective harvesting (Arndt et al., 1997; Clary et al., 2007). The manual insertion ensured that each spear was planted on the sand bed with varied angles and spacings 1-10 cm apart to replicate random growth patterns in real field conditions. Out of the total asparagus spears, 40 samples were randomly chosen as a training set for algorithm development, with the remaining 100 samples utilized as a test set to assess the camera's performance.

Three consumer-grade RGB-D cameras, operating in different depth sensing principles, were evaluated in this work, including RealSense™ L515 (Intel Corporation, Santa Clara, CA, USA), Azure Kinect (Microsoft Corporation, Redmond, WA, USA), and ZED-X Mini (Microsoft Corporation, Redmond, WA, USA), which were chosen due to their affordability and decent depth resolution necessary for precise spear localization. Technical specifications of these cameras are given in Table 1. Each camera was used independently (i.e., no data fusion among cameras), capturing both color and depth images to generate 3D point clouds for asparagus spears perception. To account for differences in the field of view among the cameras, the height of each camera was individually adjusted using a linear track, ensuring a consistent imaging field of view across all devices. Specifically, for each batch of asparagus spears, both color and depth images were acquired by each camera oriented at 0°, 25°, and 45° angles relative to the vertical axis, with the acquisition for each camera completed at all the angles before another camera was set up for imaging the same batch of spears, as shown in Figure 4. These angles were empirically chosen to examine the effect of the camera angle on spear perception and length estimation performance. A suite of Python scripts was written for image acquisition by the three RGB-D cameras, which could display the images in real-time and enable the user to trigger image acquisition via simple keystrokes. These scripts were built using the camera-specific SDK (software development kit) with a Python wrapper to interface with each camera, as shown in Table 1. As a result, a total of 72 RGB-D images were acquired across 8 batches with each of the three cameras capturing images at three different angles per batch. Each image frame consisted of a 24-bit RGB image (1980 $\times$ 1080 pixels in .png format) and a 16-bit depth map (1980 $\times$ 1080 pixels in .npy format), both spatially registered for the same scene of a specific batch of asparagus spears in the sand bin.

2.3. Field imaging

From the simulated indoor experiments described above, the ToF-based camera (Azure Kinect) was selected and configured at a fixed angle for open-field imaging, given its demonstrated performance with detailed results given in 2.2. The mobile platform with the camera was deployed for field imagery collection, as shown in Figure 5, which was conducted in different asparagus fields in the Horticultural Teaching and Research Center (Holt, MI) and the West Michigan Research Station (Hart, MI) of Michigan State University throughout the asparagus growth season (May to June) of 2024. For the ease of ground-truthing the length of the imaged spears on site, the platform was manually operated with a motor controller to traverse a row of the asparagus bed at a slow speed of about 5 cm/s during imaging acquisition. The software script used for image acquisition by the Azure Kinect in outdoor experiments was the same as that used indoors, and the format of the collected images adhered to the specifications set during indoor trials.

While the imaging platform was traveling down a specific crop row in the field, the computer screen showing the field of view was continuously

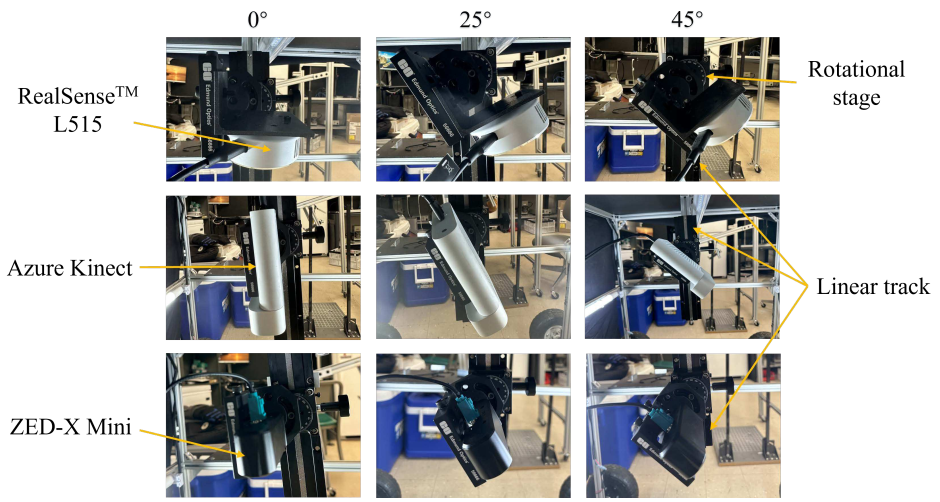

Figure 4: Photographs of three different RGB-D (red-green-blue-depth) cameras set up at varied angles with respect to the vertical axis. Each row is for the configuration of the same RGB-D camera at three different angles for asparagus imaging.

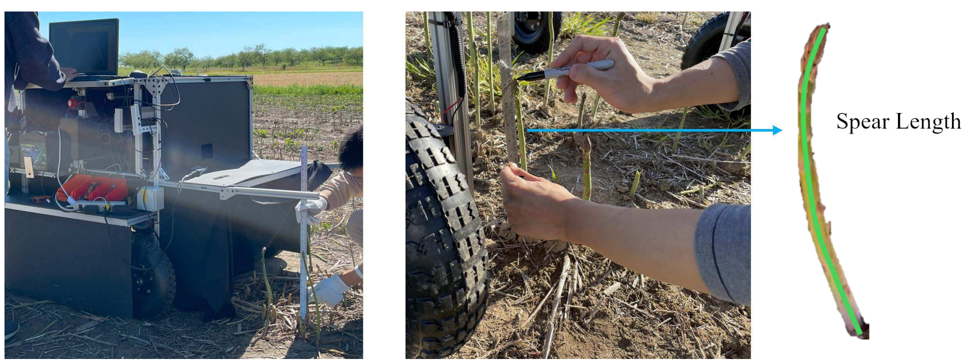

Figure 5: Photographs of three different RGB-D (red-green-blue-depth) cameras set up at varied angles with respect to the vertical axis. Each row is for the configuration of the same RGB-D camera at three different angles for asparagus imaging.

being monitored. When densely populated clusters of spears were encountered, the platform was manually controlled to stop and adjust its position slightly to capture well-centered images of the grouped spears. Right after acquiring a specific scene, the above-ground length of each asparagus spear captured in this scene was manually measured with a ruler, as shown in Fig-

Table 1: Technical specifications of three different RGB-D (red-green-blue-depth) cameras for indoor asparagus imaging. LiDAR: light detection and ranging; FOV: field of view.

Camera model	**RealSense™ L515**	**Azure Kinect**	**ZED-X Mini**
Technology	LiDAR	Time of flight	Stereo vision
Pixel resolution	1024 × 768 (depth), 1920 × 1080 (RGB)	1024 × 1024 (depth), 1920 × 1080 (RGB)	2208 × 1242 (depth per eye), 1920 × 1080 (RGB)
Field of view	70° × 55° (depth), 70° × 43° (RGB)	75° × 65° (depth), up to 120° (RGB)	Horizontal of 105° and vertical of 72° at HD1080 resolution
Depth range	0.25 m to 9 m	0.5 m to 5 m (Narrow FOV), 2.2 m to 3.8 m (Wide FOV)	High accuracy depth perception from 0.1 to 8 m (2.2 mm), 0.15 to 12 m (4 mm)
Software development kit	Librealsense SDK	Azure Kinect SDK	ZED SDK

ure 5. For each acquired image, rough annotations were created to link each measured spear to its length, helping ensure alignment between image data and measurement records. The length was defined as the full measurement from the base point of a spear on the ground to its tip end. For the curved spear, the flexible nature of the plant allowed it to be readily straightened for measurement. The length measurement is consistent with the length definition used in 3-D vision-based asparagus perception (Peebles et al., 2024; Mu et al., 2024). It should be noted that the asparagus spears shorter than a predetermined length (about 8 cm) were excluded from length measurements, as they were considered too immature and could introduce noise in the dataset. Once measurements were recorded, the platform resumed its movement until imaging was completed for the entire crop row. As a result, a total of 361 field images were acquired, alongside the length measurements of 1008 asparagus spears.

2.4. Dataset Curation

The indoor and field imaging experiments described above yielded two separate datasets. The acquired images were manually annotated using the software called Anylabeling (v0.4.8), which is an AI-assisted annotation tool that is capable of auto-labeling based on YOLOv8 (Jocher et al., 2023) and Segment Anything models (Kirillov et al., 2023). In the annotation process, as shown in Figure 6(a), those spears whose lengths were measured in the field were manually annotated with a bounding box, alongside a separate group ID assigned to the instance, and the editing tool of the annotation software was then used to add the length measurement for the asparagus spear ID.

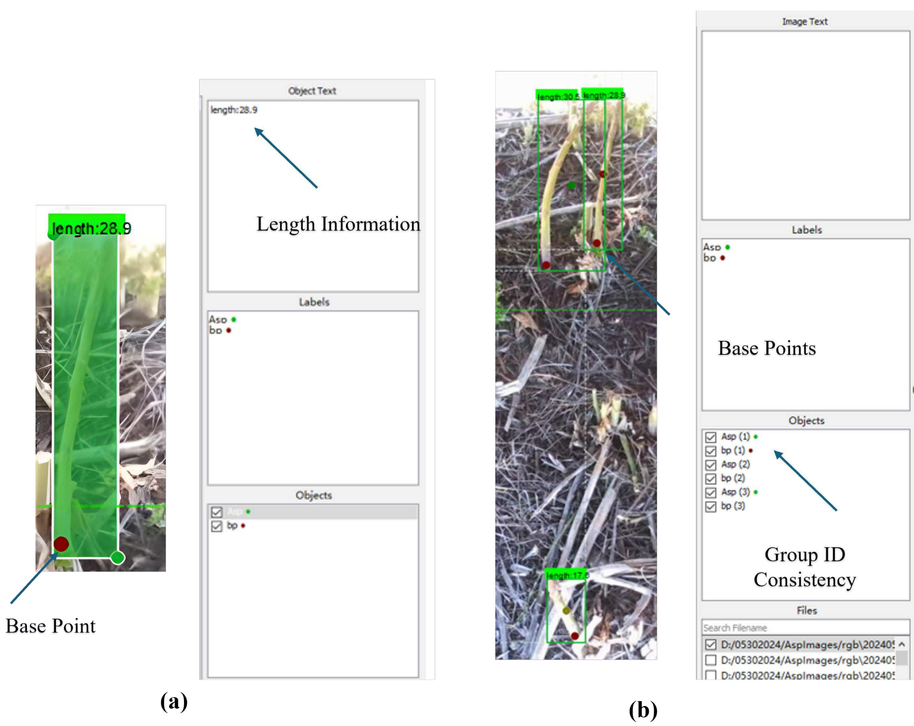

Figure 6: Asparagus spear annotation: (a) annotation of a single instance, and (b) presentation of the final image annotations.

The bounding box annotations would be used for building asparagus detection models, which were defined from the spear tip to its base at ground level. In addition, point annotations were added to mark the base location of each spear on the ground, with the same group ID as that of the corresponding bounding box annotation to create a correspondence between the base point and its bounding box, as shown in Figure 6(b). The labeled base point positions combined with the camera's spatial configuration (with the optical center defined as the origin) and the depth values obtained from the depth images, enabled us to calculate the actual base point coordinates. It is noted that the indoor imagery was only subjected to bounding box annotations since the data was only used to evaluate the camera performance in length estimation. The JavaScript Object Notation (JSON) format was initially used to store the annotation data and then converted to text (.TXT) format for training the YOLO models. Finally, the indoor data (72 images for three different cameras at three orientation angles) was annotated for 140 asparagus spears with bounding boxes, and the outdoor data (361 images by the selected camera) was annotated for 1008 spears with both bounding boxes and point annotations.

To help build a reliable model for asparagus detection, a publicly available asparagus dataset, collected by Peebles (2021) using a selective asparagus harvesting robot during 2020 field trials, was incorporated here for model development. The public dataset comprises 4359 RGB images with 17,342 bounding boxes annotations, but it lacks length measurements and depth imagery, which hence was only used to train the asparagus detection model. Consequently, an expanded dataset with 4386 RGB images was formed for model development, as shown in Figure 7, by combining 27 RGB images for 40 asparagus spears from the indoor experiment with the external dataset. These 27 images were gathered to provide consistent coverage of 40 asparagus spears, resulting in three images per spear, one from each angle (3 batches multiplied by 3 angles for each batch) to provide a solid basis for model training. The resultant dataset was then randomly split into 80% for model training and 20% for validation. The remaining 45 indoor images of 100 asparagus spears were set aside for model testing.

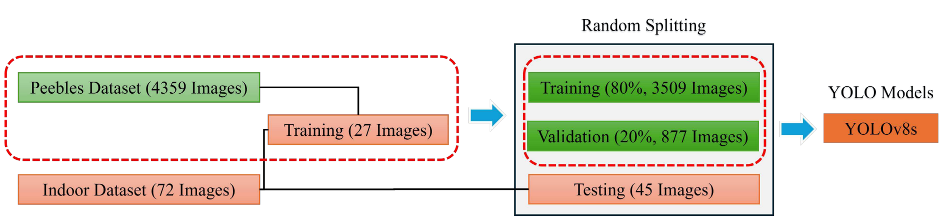

Figure 7: Flowchart of dataset curation for indoor camera performance evaluation.

To develop models for asparagus detection in the field experiment, both

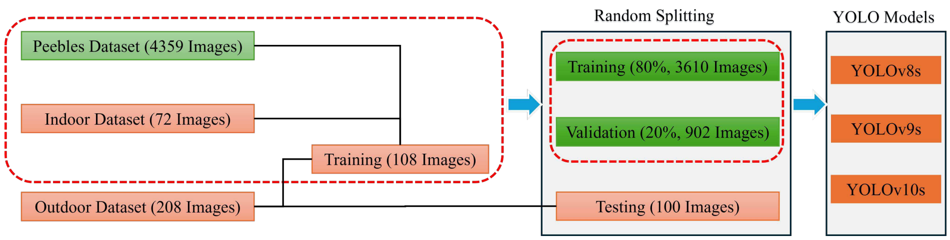

Figure 8: Flowchart of dataset curation for in-field asparagus perception validation.

the public dataset (4359 images) and the indoor image data were combined with part of the image data collected in the field in this study, as shown in Figure 8. In this way, a subset of 108 images was randomly chosen from outdoor data to assemble an expanded dataset for model development, consisting of a total of 4512 images with 19132 asparagus spears. Likewise, the dataset was randomly split into training and validation sets according to a ratio of 80% : 20%. The remaining 100 field images for 478 asparagus spears were left for model testing.

2.5. Asparagus Detection and Point Cloud Generation

In real-time object detection, the YOLO (You Only Look Once) detectors have enjoyed unprecedented popularity for their good balance of efficiency and accuracy, with 11 versions published at the time of writing. Among different YOLO models, YOLOv8 (Jocher et al., 2023), YOLOv9 (Wang et al., 2024b), and YOLOv10 (Wang et al., 2024a), which are among the forefront models that have shown respectable performance in agricultural vision tasks (Alif and Hussain, 2024), were selected for asparagus spear detection in this study. Compared to its earlier counterparts, YOLOv8 came with significant improvements in speed and accuracy by optimizing the network architecture and incorporating advanced data augmentation techniques (Jocher et al., 2023). Building upon these advancements, YOLOv9 integrated programmable gradient information and the generalized efficient layer aggregation network (Wang et al., 2024b), addressing information bottlenecks and enhancing gradient flow, thus improving detection accuracy and computational efficiency. YOLOv10 further refined these innovations with dual label assignments for non-maximum-suppression-free training, lightweight classification heads, and partial self-attention modules (Wang et al., 2024a), achieving state-of-the-art performance with lower computational costs.

Table 2: YOLO (You Only Look Once) models used in this study.

Model	Parameters (M)	URL
YOLOv8s	11.2	https://github.com/ultralytics/ultralytics
YOLOv9s	7.1	https://github.com/WongKinYiu/yolov9
YOLOv10s	7.2	https://github.com/THU-MIG/yolov10

These YOLO models are available with different scales and complexities. In this study, a small-scale variant of the three versions of YOLO detectors was selected, as shown in Table 2, for the sake of the real-time performance of asparagus detection. For the indoor evaluation of 3D cameras, only the YOLOv8s model was trained and evaluated for the task, while for the field imaginary, all three YOLO models were developed and compared to identify the best one. YOLO models were trained using the standard input image size of 640×640 pixels, with a batch size of 32, for 100 epochs, along with other default hyperparameter settings.

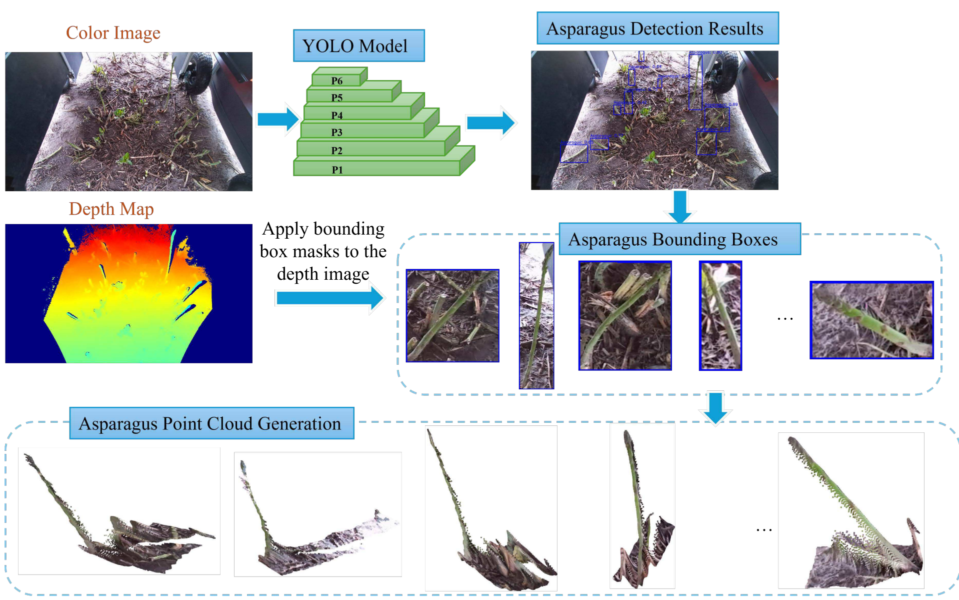

Figure 9: The proposed asparagus point cloud generation pipeline.

With an asparagus spear detected by a YOLO model, the point cloud of the spear is to be obtained for determining the spear length and base position. Figure 9 illustrates the process of asparagus detection and point cloud generation. First, the color image is processed through a YOLO model that localizes the asparagus within the image, resulting in the detection output of a series of bounding boxes for individual spears. Thereafter, these bounding boxes are applied to create masks on the depth image, isolating the depth information of each spear. The masked depth data and color image are combined to create RGB-D patches, which are then processed to generate 3-D point clouds using the known pre-calibrated intrinsic parameters of the given depth camera. The process of asparagus detection and subsequent point cloud generation was implemented using Python 3.9 with the Open3D library (version 0.18). These point clouds can be visualized as 3D models, as shown in Figure 9. It is noted that the raw point clouds generated generally contain the ground background and often suffer from noise, hindering an accurate representation of the actual asparagus spear. Hence, each asparagus point cloud requires further processing to filter out extraneous elements, as described below, before the spear length and its base position are determined.

2.6. Point Cloud Processing for Length and Base Point Prediction

After obtaining the asparagus point cloud in the initial step above, further processing was performed to segment the asparagus accurately, as illustrated in Figure 10. Since the camera may not be perfectly set up vertically in practice, the processing begins with transforming the initial point cloud to align it with the ground plane. The origin of the point cloud coordinates from a 3D camera is typically at the camera's optical center. Specifically, the point cloud was rotated around the origin by using a rotation matrix created by converting an angle of 30 degrees into radians and applying it around the x-axis as defined in Figure 10. After aligning the point cloud, the background ground plane was segmented and removed using the function *o3d.geometry.PointCloud.segment_plane()*. This function detects and removes points that belong to the largest plane in the point cloud by using the RANSAC algorithm to iteratively fit a plane model to randomly sampled points in a point cloud, with parameters such as distance threshold crucial for isolating the ground plane (Zhou et al., 2018). In this work, the distance threshold was set to 4 cm based on preliminary testing. Specifically, if a point is within 4 cm from the identified plane, it would be considered part of that plane and removed. This method distinguished the ground points from other points, removing only those very close to the ground plane, ensuring accurate isolation of individual asparagus spears from the ground in the point cloud.

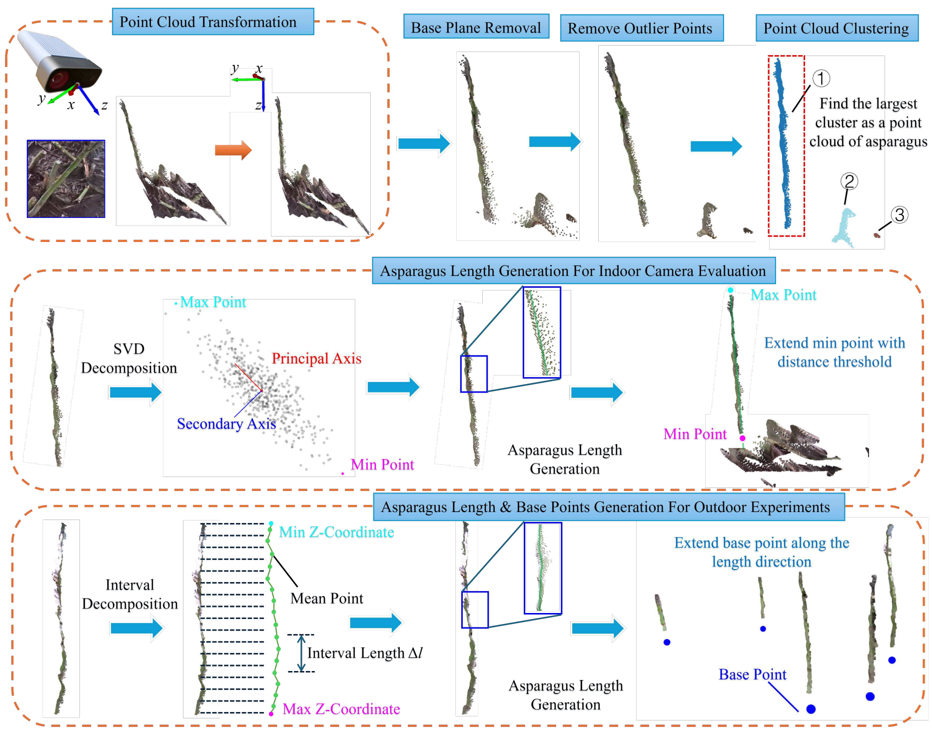

Figure 10: Asparagus point cloud processing pipeline. SVD denotes singular value decomposition.

With the ground plane removed, a further operation was to filter out outlier noise by using *o3d.geometry.PointCloud.remove_statistical_outliers()* function, which examines neighbors of each point (30 neighbors in this case) and removes points that are further than a certain standard deviation (empirically set to 1.0 here) from the average distance to their neighbors, thus refining the spear point cloud by selecting only the points that meet the criterion. Unlike Peebles et al. (2020), the refined point cloud further underwent density-based clustering using DBSCAN, implemented using the function *3d.geometry.PointCloud.cluster_dbscan()*, to accurately identify individual asparagus spears, particularly in densely clustered scenarios. The DBSCAN algorithm clusters point cloud data by grouping points that are closely packed together based on a specified distance (empirically set to 2 cm in this case) and a minimum point threshold of 10 to identify and segment individual asparagus spears. Among these, the largest cluster, representing the primary asparagus spear, was isolated as the final asparagus point cloud to provide a clear and focused dataset for further analysis.

For 3D camera evaluation, PCA (Principal Component Analysis) was applied to estimate the asparagus spear length from its isolated point cloud, because of the ease and decent accuracy of this approach (Abdi and Williams, 2010). However, in field conditions where asparagus grows at various angles and often is curved due to its growth nature, the PCA-based approach would be inappropriate for length estimation. Instead, to estimate the length and base point of an asparagus spear from its point cloud, the points were segmented along the Z-axis into thin slices based on a defined segment height Δl (set to 1 cm here), as shown in Figure 10. The mean point of each segment was then determined to reduce noise and summarize the spear's position at each height. The mean point list was used to form a connected piecewise line segment representing the spear, and then, the length of the spear was estimated by the summation of the Euclidean distance between consecutive mean points. One should note that after removing the ground plane, parts of the asparagus point cloud might be removed. Therefore, the actual length of the asparagus should include the distance from the lowest point in the mean point list of the asparagus point cloud to the base point. The base point was determined by extending the line connecting the last two points in the mean point list down to the soil plane by an extended distance using a distance threshold derived using the function *o3d.geometry.PointCloud.segment_plane()*. This method ensures any missing parts of the asparagus near the base would be accounted for in the length calculation.

In this work, both model development for asparagus detection and point cloud processing were conducted on a computer with an Intel Core (TM) i9-11900K @ 3.50 GHz processor with 64GB RAM and NVIDIA GeForce RTX 3060 GPU (12 GB memory) in Windows 11 OS.

2.7. Performance Metrics

A set of metrics was employed to assess the performance of the proposed methodologies for vision tasks on the training and testing data, including asparagus model training, detection, length estimation, and base point prediction. For the asparagus detection model, the mean average precision at an intersection over union (IoU) threshold of 0.5 (mAP50) was used, which measured the accuracy of detecting objects and predicting their bounding boxes:

$$\mathrm{mAP50} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{AP}_i \tag{1}$$

Where N represents the total number of classes. For asparagus detection, as done in previous studies (Peebles et al., 2019; Mu et al., 2024), the metrics including *Precision*, *Recall*, and *F1 score*, are calculated as follows:

$$\text{Precision} = \frac{\#TP}{\#TP + \#FP} \tag{2}$$

$$\text{Recall} = \frac{\#TP}{\#TP + \#FN} \tag{3}$$

$$\mathit{F1\ score} = 2 \times \frac{\mathit{Precision} \times \mathit{Recall}}{\mathit{Precision} + \mathit{Recall}} \tag{4}$$

Where $\#TP$, $\#FP$, and $\#FN$ represent the number of true positives, false positives, and false negatives, respectively. In the test image, TP indicated successfully identified and located green asparagus (defined by IoU values greater than 0.5 for a given detection), FP referred to non-green targets mistakenly recognized as asparagus, and FP indicated instances where green asparagus was not successfully identified. Precision measures the proportion of correctly identified asparagus spears (i.e., TP) among all detected spears, including FP, i.e., the non-spears incorrectly identified as spears. Recall reveals how well the detection model identifies the spears among all the actual spears, including FP, i.e., these spears not detected by the model. The *F1 score* is a balanced metric between *Precision* and *Recall*, giving a comprehensive assessment of the performance of detection models.

The asparagus spear length estimation was assessed using the metrics including root mean square error ($RMSE$), coefficient of determination (R^2), and percent accuracy. These metrics were calculated for spears in the test set by comparing the measured length and predicted value, as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5}$$

$$\mathit{Percent\ Accuracy} = \frac{1}{n}\left(\sum_{i=1}^{n}\left(1 - \frac{|y_i - \hat{y}_i|}{y_i}\right)\right) \times 100\% \tag{6}$$

Where y_i, $\hat{y}_i$, and $\bar{y}_i$ are the measured length, predicted length, and the mean of the measured lengths of all spears in the test set, respectively, and n represents the total number of asparagus spears in the test set. Furthermore, to evaluate the accuracy of base point predictions, in addition to using

RMSE along the x, y, and z axes, the Euclidean distance (E_d) error was calculated as follows:

$$E_d = \sqrt{(x_{\text{pred}} - x_{\text{actual}})^2 + (y_{\text{pred}} - y_{\text{actual}})^2 + (z_{\text{pred}} - z_{\text{actual}})^2} \quad (7)$$

Where $x_{\text{pred}}, y_{\text{pred}}, z_{\text{pred}}$ are the predicted coordinates of the base point of a given spear in 3D space, and $x_{\text{actual}}, y_{\text{actual}}, z_{\text{actual}}$ are the corresponding actual coordinates of the base point. The E_d offers a measure of the localization accuracy of the spear base point in 3D space. Lower E_d values indicate greater accuracy in base point localization, which is important for spear length estimation and precision cutting of eligible spears.

3. Results & Discussion

3.1. Indoor Evaluation of Camera Performance in Asparagus Perception

Table 3 summarizes the asparagus detection results by three 3-D cameras with YOLOv8s. Here, a test set of 100 asparagus spears was used for performance evaluation of the cameras oriented at three different angles (Figure 4). Overall, the Azure Kinect appeared to perform the best, followed by the ZED-X Mini and RealSense$^{\text{TM}}$ L515; at 45°, the Azure Kinect achieved the highest *F1 score* of 0.903 on par with the ZED-X Mini. Although the three cameras captured images at the same resolution, their differences in sensor quality, lens characteristics, and noise levels may explain the different performances in asparagus detection. The results also reveal that the detection accuracy at 25° and 45° were generally higher than that at 0° for all cameras. At wider angles, the improved visibility and separation of the spears enabled the cameras to recognize better and delineate individual spears, increasing detection accuracy. However, the RealSense$^{\text{TM}}$ L515 struggled more at 45° with its *F1 score* down to 0.833. Despite no extensive examination of camera orientation, these results indicate the angle substantially affected the asparagus detection, with 25° and 45° advantageous over 0°. The results of indoor asparagus length estimation by the three cameras are summarized in Table 4. Since the point cloud processing and length estimation were based on the initial detection of the actual asparagus spears, only spears correctly detected by YOLOv8s, i.e., TPs, were included in the following analysis for length estimation. The Kinect Azure remained the top performer, followed by the ZED-X Min and then RealSense$^{\text{TM}}$ L515. For both the Kinect Azure and the ZED-X Min, the accuracy peaked at 25°, while the RealSenseTM L515 gave the best accuracy at 45°; the performance difference between 25°

Table 3: Indoor asparagus detection by three red-green-blue-depth cameras

Camera	Angle	TP	FN	FP	Precision	Recall	F1 Score
RealSense™ L515	0°	83	17	18	0.826	0.831	0.828
	25°	88	12	15	0.859	0.879	0.869
	45°	82	18	15	0.848	0.818	0.833
Kinect Azure	0°	89	11	13	0.871	0.894	**0.882**
	25°	91	9	11	0.893	0.912	**0.902**
	45°	93	7	12	0.882	0.925	**0.903**
ZED-X Mini	0°	88	12	15	0.853	0.881	0.867
	25°	90	10	13	0.874	0.903	0.888
	45°	91	9	11	0.892	0.914	0.903

and 45° was relatively marginal for all three cameras. The overall performance trend is consistent with that in asparagus detection where the Kinect Azure produced the best accuracy.

Table 4: 3D Camera Length Estimation Performance

Camera	Angle	R^2	RMSE	Accuracy
RealSense™ L515	0°	0.713	3.9	83.1%
	25°	0.826	2.8	87.3%
	45°	0.828	2.7	87.1%
Kinect Azure	0°	0.785	3.8	83.9%
	25°	0.864	2.5	89.0%
	45°	0.862	2.5	89.2%
ZED-X Mini	0°	0.746	3.7	83.2%
	25°	0.849	2.6	87.8%
	45°	0.837	2.5	86.4%

The angled positioning of cameras (e.g., at 25° and 45°) allowed a better longitudinal view of asparagus spears in the imaged scene, reducing occlusions that would occur when the camera was positioned directly above the spears. Even though occlusions can be minimal at 0° since objects are seen directly from above, these views can miss important side details and depth variations that angled views otherwise would capture, especially for elongated objects like asparagus spears. Nevertheless, positioning the camera at an even steeper angle (e.g., at 60°) may lead to performance drops due to the inherent challenges associated with occlusion issues that become pronounced as the camera captures more of the side profile of spears, and the bases of the spears could not be visible. The lower parts (e.g., base points) of the spears, especially where they meet the ground, can become obscured or cut off due to the steep perspective, negatively impacting the accuracy of length measurements. In both the spear detection and length estimation tasks, the Kinect Azure delivered the best performance across all tested angles. Given the indoor experiment findings, the Kinect Azure angled at a fine-tuned angle of 30° was implemented for field imagery acquisition to ensure the visibility of spear bases while minimizing occlusions.

3.2. Performance of Field Experimentation

3.2.1. Reference measurements

Figure 11 illustrates the distribution of a total of 1008 length measurements of asparagus spears conducted during imaging in the field. The spear length ranged from approximately 5 cm (2 inches) to 50 cm (20 inches), covering a diversity of maturity levels. The distribution was skewed towards shorter lengths, with 44.6% of spears falling between 15 cm (6 inches) and 25 cm (10 inches), and the counts of spears decreased progressively for those longer than 25 cm, with very few spears exceeding 40 cm. Such distribution could provide meaningful variations for testing detection models and asparagus perception pipeline, demonstrating their reliability and ability to handle different spear sizes.

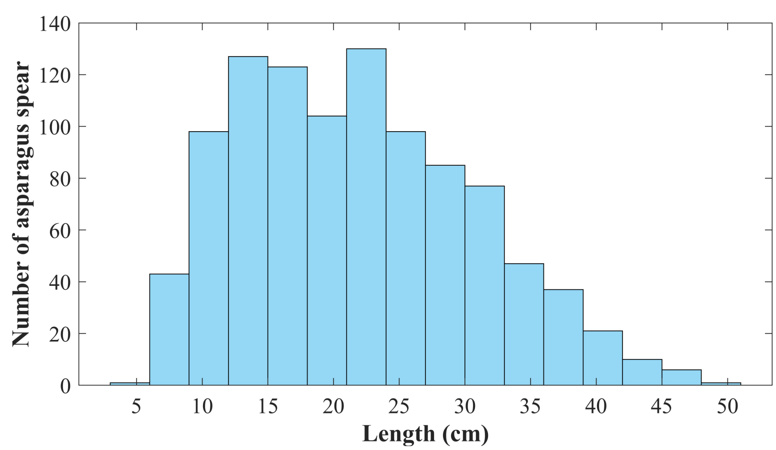

Figure 11: Asparagus spear length distribution.

Figure 12 presents two scatter plots illustrating the spatial distribution of asparagus spear base points in physical 3-D space, with the origin of the point cloud coordinate at the camera's optical center. As shown in Figure 12(a), the base points ranged from approximately -80 cm to 60 cm along the X-axis and from -60 cm to 40 cm along the Y-axis. The distribution appeared to be relatively dense near the center with a gradual dispersion toward the boundary, which is consistent with previous the field of view setting when acquiring images. The variation in spatial distribution ensures that the data covers a range of conditions, from areas where the camera was positioned closer for detailed views to those further away for broader coverage, which is essential for evaluating the performance of the developed pipeline across different perspectives and field conditions. Regarding the base points along the Z-coordinate (height), their height ranged from 85 cm to 125 cm with 86.7% between 95 cm and 115 cm, as shown in Figure 12(b). Although the camera was set to about 1 m from the ground, the Z coordinate fluctuated because the field terrain was uneven in general.

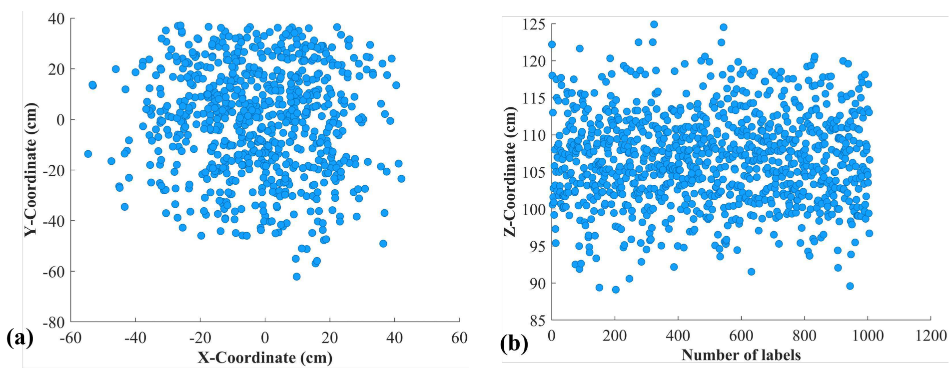

Figure 12: The asparagus spear base point coordinates distribution (with the origin of the point cloud coordinate defined at the camera's optical center).

3.2.2. Asparagus Detection

Figure 13 presents mAP50 profiles during the training of three YOLO models for asparagus spear detection. The training performance of the three models increased rapidly within the first 20 epochs and appeared to level off with minor fluctuations after around 40 epochs. Despite initial disparities in model training performance, YOLOv8s demonstrated the best training accuracy reaching its peak around the 75th epoch with an mAP50 of 82.1%. This was followed by YOLOv10s, which is slightly below YOLOv8s, especially after 50 epochs. YOLOv9s initially showed strong performance during the early 30 epochs. However, after this point, YOLOv9s progress plateaued and its mAP50 stabilized at a lower value compared to the other models.

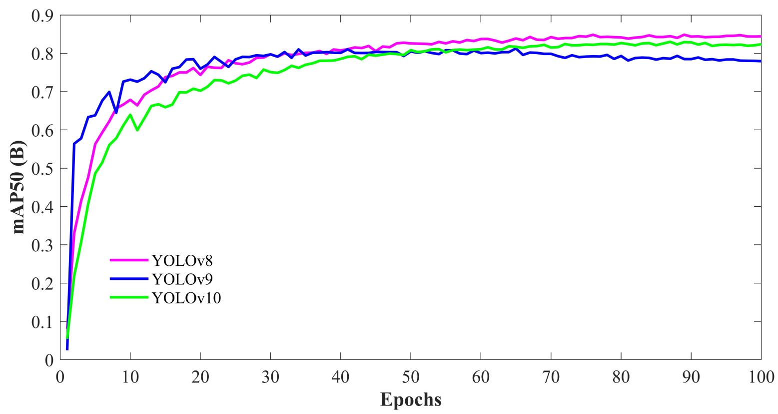

Figure 13: Training curves of three YOLO models for asparagus detection.

Table 5 summarizes the testing accuracy of the three models. YOLOv8s achieved the best performance, outperforming the other two models by a larger margin, with *Precision* = 0.866, *Recall* = 0.832, and *F1 Score* = 0.849, indicating its ability to accurately detect asparagus spears while minimizing both false positives and false negatives. Likewise, YOLOv10s, with an F1 Score of 0.828, was second to YOLOv8s, followed by YOLOv9s with the lowest F1 Score of 0.809. Given the best asparagus detection accuracy attained by YOLOv8s, this model was selected in the subsequent analyses to ensure optimal performance.

Table 5: YOLO models testing results. #TP, #FN, and #FP denote the number of true positives, false negatives, and false positives, respectively.

Models	#TP	#FN	#FP	Precision	Recall	F1 Score
YOLOv8s	398	80	61	**0.866**	**0.832**	**0.849**
YOLOv9s	381	97	82	0.823	0.796	0.809
YOLOv10s	391	87	76	0.837	0.818	0.828

3.2.3. Spear Length Estimation and Base Points Prediction

Figure 14 presents a scatter plot of the predicted versus measured lengths of asparagus spears, where only correctly detected spears (e.g., TP of YOLOv8s) are included in this analysis. The data points cluster closely around the diagonal line representing the agreement between predicted and true lengths.

The coefficient of determination of $R^2 = 0.918$ suggested a strong correlation between the predicted and actual values and hence the validity of the developed algorithm pipeline in estimating spear lengths. The respectable performance is also reflected in the $RMSE$ of 2.316 cm (less than 1 inch) and the overall percentage accuracy of 89.4%.

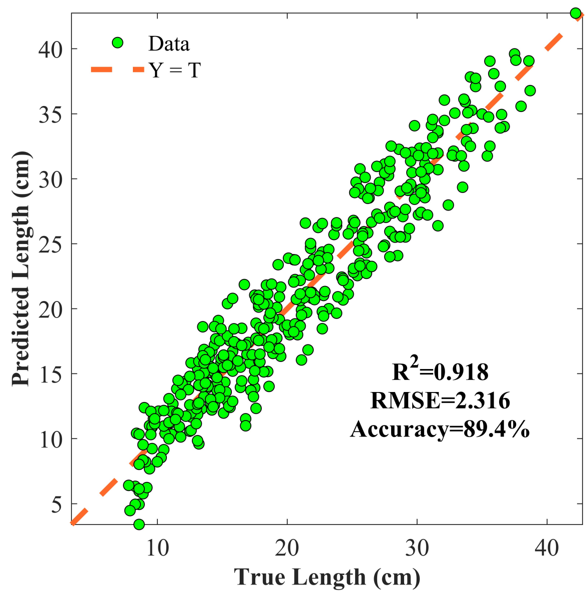

Figure 14: Asparagus spears length estimation results.

Figure 15 compares the predicted and true coordinates (X, Y, and Z) of the base points of asparagus spears. Higher accuracies were obtained in predicting the X and Y coordinates than in the Z coordinates. Nevertheless, the $RMSE$ values in predicting all three coordinates were smaller than 2 cm, indicating the efficacy of the developed pipeline for localizing the spear base points. This level of localization accuracy and even better is essential for ensuring that appropriately designed selective harvesting mechanisms can precisely target the spears without damaging neighboring spears.

Figure 16 shows the distribution of Ed errors in predicting the position of asparagus spear base points in 3D space. The majority (51.4%) of the localization errors are concentrated between 1 cm and 2 cm, with the most frequent errors around 1.5 cm and the mean error of 1.4 cm. The distribution tails off with 85.9% of spears predicted with errors less than 2 cm, and only 1.2% have errors exceeding 3 cm. This distribution suggests that the base point localization by the proposed method was overall accurate, except that

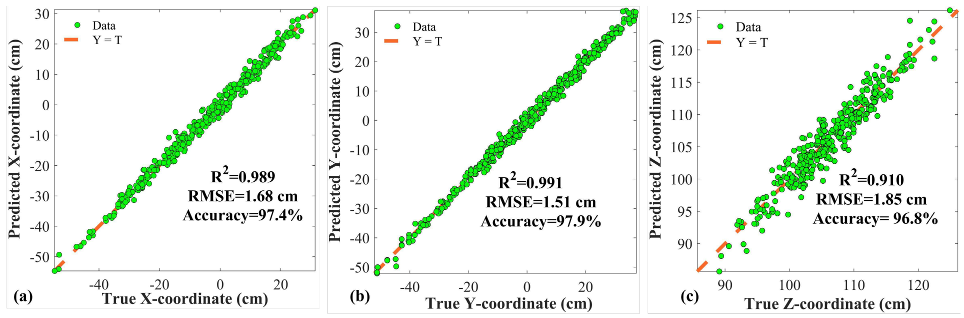

Figure 15: Base points estimation results for three coordinates (X, Y, Z). It is noted that the origin of the point cloud coordinate is defined at the camera's optical center).

occasional outliers existed with the predictions deviating significantly from the actual position.

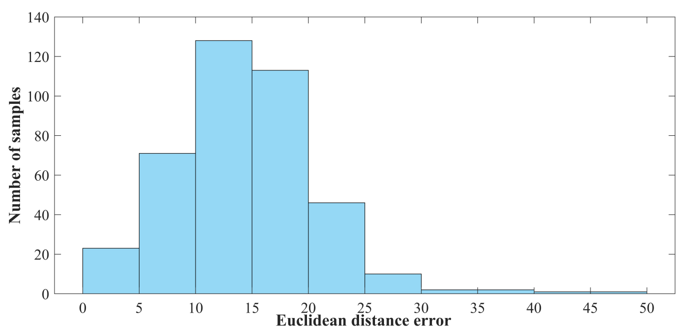

Figure 16: The distribution of Euclidean distance (E_d) errors.

3.2.4. Harvest Eligible Spear Analysis

In selective harvesting of asparagus, spears longer than 7-9 inches are generally considered ready for harvesting in the U.S. (Read, 2009). To effectively harvest the eligible spears by harvesting mechanisms, precise localization of spear base points is crucial because even slight deviations can result in either harvest failure or collateral damage to nearby spears, leading to reduced yield and quality. Hence, to determine the harvest eligibility of detected spears, the following two conditions, ranked by importance, were adopted in this study:

- **Condition 1**: Detected spears must have a predicted length greater than 8 inches (20.3 cm).

- **Condition 2**: The E_d error of the detected spear base point must be less than 2 cm.

Tables 6 and 7 present the confusion matrices of detected spears that are categorized based on their harvest eligibility under the two conditions. With Condition 1 applied alone, the developed pipeline predicted 95.8% of (186 of the 194) actual harvestable spears, with only 8 harvestable spears incorrectly predicted as not harvestable. On the other hand, the developed pipeline achieved 95.7% accuracy in identifying non-harvestable spears. With both Conditions 1 and 2 applied (Table 7), the accuracy for predicting harvestable spears dropped to 84.5%, with 164 of the 194 actual harvestable spears correctly identified. The number of false negatives (eligible spears miscategorized as ineligible) increased to 30, indicating that some spears that met Condition 1 (the length criterion) were excluded due to localization errors of spear base points. Conversely, the accuracy for non-harvestable spears improved slightly to 97.1%, with only 6 spears incorrectly classified as harvestable.

Table 6: Confusion matrix of detected spears with Condition 1 alone applied.

Only with Condition 1	Predicted Harvestable	Predicted Not Harvestable
Actual Harvestable	**186 (95.8%)**	9
Actual Not Harvestable	8	**203 (95.7%)**

Table 7: Confusion matrix of detected spears with both Conditions 1 and 2 applied.

With Conditions 1 and 2	Predicted Harvestable	Predicted Not Harvestable
Actual Harvestable	**164 (84.5%)**	30
Actual Not Harvestable	6	**206 (97.1%)**

Tables 8 and 9 present the harvest eligibility analysis of all ground-truth spears in the test set, including all detected and missed spears. When only Condition 1 was applied (Table 8). the developed pipeline correctly identified 186 out of 231 actual harvestable spears with 80.5% accuracy. However, 45 harvestable spears were incorrectly classified as not harvestable, indicating a significant number of false negatives. For non-harvestable spears, the developed pipeline yielded 76.0% accuracy, with 55 false positives (non-harvestable spears misclassified as harvestable). With both Conditions 1 and

2 met (Table 9), the accuracy for predicting harvestable spears decreased to 71.0%, with the number of false negatives up to 67. Apparently, compared with the analysis for only detected spears, as shown in Tables 6 and 7, the accuracy decreased substantially when all ground-truth spears were considered. The inclusion of Condition 2 reduced the accuracy of harvest-eligible spears but increased the accuracy of non-harvestable spears, suggesting that the developed pipeline became more conservative and accurate in avoiding false positives at the expense of missing some spears that met the length requirement.

Table 8: Confusion matrix of all ground-truth spears with Condition 1 applied.

Only with Condition 1	Predicted Harvestable	Predicted Not Harvestable
Actual Harvestable	**186 (80.5%)**	45
Actual Not Harvestable	55	**203 (76.0%)**

Table 9: Confusion matrix of all ground-truth spears with both Conditions 1 and 2 applied.

With Conditions 1 and 2	Predicted Harvestable	Predicted Not Harvestable
Actual Harvestable	**164 (71.0%)**	67
Actual Not Harvestable	6	**203 (95.7%)**

3.2.5. Processing Time Analysis

Table 10 provides a detailed breakdown of the processing time of the developed asparagus perception pipeline with YOLOv8s. The time was obtained by running the pipeline on the onboard computer (Jetson AGX Orin^{TM} 64-GB Developer Kit, NVIDIA) and averaging the results over 10 test frames. The YOLOv8s took 38 ms for asparagus spear detection. The point cloud generation tasks including depth map masking and point cloud creation, were faster with 32 ms. However, RANSAC-based ground removal and subsequent statistical outlier removal took 61 and 43 ms, contributing to over a third of the overall processing time. The most time-consuming single step was DBSCAN clustering which took 79 ms. The density-based approach for clustering could slow down processing in dealing with large sets of point clouds. Finally, the length estimation and base point estimation steps took 34 ms and 17 ms, respectively.

Table 10: Processing time breakdown of the developed asparagus perception pipeline.

Steps	Time (ms)
Asparagus spear detection by YOLOv8s	38
Masking depth maps based on bounding boxes	6
Generating the point cloud	26
Initial point cloud filtering (removing ground plane)	61
Statistical outlier removal	43
DBSCAN clustering	79
Length estimation (slicing and calculating mean points)	34
Spear base point estimation	17
Total processing time	**304**

With a total processing time of 304 ms, the proposed pipeline was capable of processing image streams at about 3.3 FPS (frame per second), meeting the need for real-time asparagus proception. There are several ways to enhance processing efficiency, especially in point cloud processing, such as optimizing the clustering algorithm by using parallelized versions of DBSCAN (Götz et al., 2015; McInnes et al., 2017). The implementation of the pipeline in C++, rather than in Python in this work, would substantially accelerate the entire process. Downsampling the generated point cloud can speed up downstream computational tasks including outlier removal and clustering. However, caution should be exerted in performing downsampling since reducing data points aggressively could comprise perception accuracy, particularly for spear length estimation. Using TensorRT can accelerate the asparagus detection process by optimizing the neural network model (Jeong et al., 2022), although this has no major effects on the entire pipeline. Moreover, upgrading computing hardware would offer a quick solution for enhanced real-time efficiency without compromising the portability of the mobile vision system.

3.3. Discussion

Effective asparagus perception is necessary for enabling selective spear harvesting. This study presents a 3D vision system with a computer algorithms pipeline for detecting and localizing asparagus spears in the field. To the best of our knowledge, this represents the first systematic research into 3D vision for asparagus perception toward developing selective harvesting technology of asparagus in the U.S. A comprehensive evaluation of different

3D imaging sensors, which had not been explored in previous research, was conducted in this study, and the best vision sensor was determined for subsequent field imagery, alongside with a large set of ground-truth spear length measurements conducted in the field. The developed algorithm pipeline has yielded encouraging perception accuracy in spear detection, length estimation, and base point localization, with real-time efficiency.

Based on the indoor 3D camera evaluation, the ToF-based Azure Kinect was identified as the most suitable camera for asparagus perception in an angled configuration. The efficacy of ToF imaging sensors for asparagus perception has been shown in earlier studies (Leu et al., 2017; Peebles et al., 2020), where the Microsoft Kinect v2 was used in a selective asparagus harvesting robot. However, the earlier generation ToF camera, which was leased in 2014, has a noticeably lower depth map resolution of 512 $\times$ 424 pixels, in addition to its lower color image resolution, compared to its successor, i.e., the Azure Kinect in this study (Kurillo et al., 2022). The latter offers a depth resolution of up to 1024 $\times$ 1024 pixels, providing enhanced depth precision and adaptability as demonstrated in different studies (Kurillo et al., 2022; Neupane et al., 2021). The Azure Kinect was discontinued by Microsoft in 2023, but the same imaging technology remains available on the market through several partner options (Mehta, 2023). In addition to the 3D cameras examined in this study, there are competitive options available, as 3D vision technology advances rapidly.

Among the three small-scale yet advanced YOLO (v8, v9, and v10) models for asparagus detection, the YOLOv8s achieved an *F1 score* of 0.849, which compared favorably to those reported in previous work. Peebles et al. (2019) applied the Faster R-CNN model to a similar green asparagus dataset and achieved a lower *F1 score* of 0.73. Some studies reported higher accuracy in asparagus detection, such as the YOLOv5-spear in Zhang et al. (2022) with an F1 score of 0.975, and the HGCA-YOLO in Zhang2024 with mAP exceeding 95%, but it should be noted that these studies were focused on white asparagus grown in protected facilities (e.g., greenhouse) conditions. For enhanced 3D localization of green asparagus, Mu et al. (2024) proposed the YOLOv8-based S2CPL model that achieved an IoU of 98% for asparagus segmentation and a subsoil cutting-point accuracy of over 99%. Likewise, it is crucial to consider that this work was conducted in a greenhouse with much more structured environments than open field conditions. In the U.S., green asparagus is cultivated outdoors where the random growth pattern with clusters of spears and unstructured field conditions (e.g., variable lighting, soil background, and the presence of harvested spears left on the ground) present a multitude of challenges to asparagus perception.

For spear length estimation, this study created a new annotated imagery dataset of 1008 asparagus spears alongside length measurements manually conducted on-site in diverse field conditions. The dataset volume exceeds by far the spear length data in Mu et al. (2024) where only 25 asparagus spears were measured indoors for performance assessment, providing a more solid testbed for algorithm evaluation. In identifying harvest-eligible spears in terms of the predicted length of detected spears, this study achieved a success rate of 95.8%, outperforming the success rate of 92.3% in field tests reported byPeebles et al. (2019) and the 90% success rate in Leu et al. (2017), althogh such comparison here is not based on the same database. However, the success rate decreased to 84.5%, when high base point localization accuracy (with Ed error lower than 2 cm) was required, which is desired in the design of harvesting end-effectors selectively targeting eligible spears with minimal collateral damage. Nevertheless, the localization accuracy of the vast majority (approximately 98%) of spears remained within 3 cm (Figure 12).

There is still room for improved asparagus detection. In addition to expanding the asparagus dataset in future growth seasons, more advanced detection models, such as YOLOv11 and real-time vision transformers (Lv et al., 2024; Zhao et al., 2024) will be examined to enhance spear detection accuracy. Future efforts will also focus on optimizing the entire algorithm pipeline to make it more computationally efficient without compromising detection and localization accuracy. Data curation efforts will be tailored to the specific challenges of asparagus detection to evaluate model generalization performance across changing field conditions. The developed 3D vision system of this study has yet to be integrated with proper harvesting mechanisms to develop and evaluate a new automated, selective asparagus harvester. Dedicated research is underway to design end-effectors capable of accurate and efficient spear harvesting specifically for asparagus in the U.S.

4. Conclusion

This study presents the development and evaluation of a 3D vision-based perception system of green asparagus for selective harvesting. Indoor evaluation of three 3D imagers shows that the ToF-based camera outperformed the two other LiDAR and stereovision-based cameras in asparagus detection and length estimation. A mobile vision system with the selected 3D camera based on indoor evaluation was deployed for field imagery collection and curation of asparagus alongside ground-truth length measurements of individual spears in diverse field conditions. The developed asparagus perception pipeline with advanced YOLO detectors performed a series of tasks including asparagus detection, length estimation, and localization of spear base points. Among the three small-scale YOLO models (v8, v9, and v10), YOLOv8s achieved the best F1 score of 0.849. By processing segmented point clouds of detected spears, the spear length estimation accuracy of 89.4% was achieved, with the corresponding RMSE of 2.316 cm, and the averaged location error for spear base points was 1.4 cm, with 85.9% of spears localized with the error less than 2 cm. The harvest eligibility analysis of detected spears showed a detection rate of 95.8% of harvestable spears when only the length criterion was applied and a detection rate of 84.5% when both the length and base point localization criteria were considered. The developed pipeline showed good promise for onboard, real-time asparagus perception. Future work will focus on integrating the 3D vision system with harvest mechanisms for selective harvesting while optimizing the pipeline for improved perception and faster processing.

Acknowledgment

This work was in part supported by the Michigan Vegetable Council and the Hatch Multistate Research Fund from U.S. Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA). The authors thank Boyang Deng and Mingjun Li for assisting in-field asparagus imagery collection and Jiaming Zhang and Parvin Mohammadi for image annotation.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2, 433–459. doi:10.1002/wics.101.

Alif, M.A.R., Hussain, M., 2024. Yolov1 to yolov10: A comprehensive review of yolo variants and their application in the agricultural domain. arXiv preprint arXiv:2406.10139 doi:10.48550/arXiv.2406.10139.

Arndt, G., Rudziejewski, R., Stewart, V., 1997. On the future of automated selective asparagus harvesting technology. Computers and Electronics in Agriculture 16, 137–145. doi:10.1016/S0168-1699(96)00033-6.

Bardenhagen, C., Werling, B., Colella, F., 2023. Michigan asparagus cost of production study 2022-23. Michigan State University Extension Bulletin E3483. URL: https://www.canr.msu.edu/vegetables/uploads/files/MIAsparagusCost_6.23_AA.pdf.

Baylou, P., Amor, B.E.H., Monsion, M., Bouvet, C., Boussau, G., 1984. Detection and three-dimensional localization by stereoscopic visual sensor and its application to a robot for picking asparagus. Pattern Recognition 17, 377–384. doi:10.1016/0031-3203(84)90067-0.

Cembali, T., Folwell, R.J., Ball, T., Clary, C.D., 2005. Economic comparison of selective and non-selective mechanical harvesting of asparagus, in: 2005 ASAE Annual Meeting, p. Paper No. 053003. doi:10.13031/2013.19053.

Chen, D., Zhang, Q., Wang, S., 2010. Current status and technical challenges of asparagus mechanical harvesting, in: 2010 Pittsburgh, Pennsylvania, June 20-June 23, 2010. doi:10.13031/2013.29743.

Clary, C., Ball, T., Ward, E., Fuchs, S., Durfey, J., Cavalieri, R., Folwell, R., 2007. Performance and economic analysis of a selective asparagus harvester. Applied Engineering in Agriculture 23, 571–577. doi:10.13031/2013.23665.

Drouin, M.A., Seoud, L., 2020. Consumer-grade rgb-d cameras, in: Liu, Y., Pears, N., Rosin, P.L., Huber, P. (Eds.), 3D Imaging, Analysis and Applications. Springer, Cham, pp. 215–264. doi:10.1007/978-3-030-44070-1_5.

FAOSTAT, 2024. Value of agricultural production. Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/faostat. accessed on November 10, 2024.

Grattoni, P., Cumani, A., Guiducci, A., Pettiti, G., 1994. Automatic harvesting of asparagus: an application of robot vision to agriculture, in: Mobile Robots VIII, pp. 200–210. doi:10.1117/12.167496.

Grattoni, P., Guiducci, A., 1990. Contour coding for image description. Pattern Recognition Letters 11, 95–105. doi:10.1016/0167-8655(90)90119-M.

Götz, M., Bodenstein, C., Riedel, M., 2015. Hpdbscan: Highly parallel dbscan, in: Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments, pp. 1–10. doi:10.1145/2834892.2834894.

Haws, S.K., 1997. Stalk selective harvesting machine. URL: https://patents.google.com/patent/US5501067A/en.

Haws, S.K., 2018. Selective harvester. URL: https://patents.google.com/patent/US8656695B2/en.

Herrygers, C., 2024. The clock is ticking for specialty crop growers. Focus on Agriculture. URL: https://www.fb.org/focus-on-agriculture/the-clock-is-ticking-for-specialty-crop-growers. accessed on November 10, 2024.

Hong, W., Ma, Z., Ye, B., Yu, G., Tang, T., Zheng, M., 2023. Detection of green asparagus in complex environments based on the improved yolov5 algorithm. Sensors 23, 1562. doi:10.3390/s23031562.

Humburg, D., 1991. The application of machine vision to the selective harvest of green asparagus. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Urbana, IL, USA. URL: https://www.ideals.illinois.edu/items/23737.

Humburg, D.S., Reid, J.F., 1991. Field performance of machine vision for the selective harvest of asparagus. SAE transactions , 81–92doi:10.4271/911751.

Irie, N., Taguchi, N., 2014. Asparagus harvesting robot. Journal of Robotics and Mechatronics 26, 267–268. doi:10.20965/jrm.2014.p0267.

Irie, N., Taguchi, N., Horie, T., Ishimatsu, T., 2009. Asparagus harvesting robot coordinated with 3-d vision sensor, in: 2009 IEEE International Conference on Industrial Technology, pp. 1–6. doi:10.1109/ICIT.2009.4939556.

Jeong, E., Kim, J., Ha, S., 2022. Tensorrt-based framework and optimization methodology for deep learning inference on jetson boards. ACM Transactions on Embedded Computing Systems (TECS) 21, 1–26. doi:10.1145/3508391.

Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics yolo (version 8.0.0) [computer software]. URL: https://github.com/ultralytics/ultralytics.

Kepner, R., 1959. Mechanical harvester for green asparagus. Transactions of the ASAE 2, 0084–0087. doi:10.13031/2013.41176.

Kepner, R., 1971. Selective versus nonselective mechanical harvesting of green asparagus. Transactions of the ASAE 14, 405–0410. doi:10.13031/2013.38301.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026. doi:10.48550/arXiv.2304.02643.

Knicely, D.R., 1962. Harvesting asparagus mechanically; economic aspects and physical properties. Master Thesis, Michigan State University, East Lansing, MI, USA. URL: https://search.worldcat.org/title/19658363.

Kurillo, G., Hemingway, E., Cheng, M.L., Cheng, L., 2022. Evaluating the accuracy of the azure kinect and kinect v2. Sensors 22, 2469. doi:10.3390/s22072469.

Leu, A., Razavi, M., Langstädtler, L., Ristić-Durrant, D., Raffel, H., Schenck, C., Gräser, A., Kuhfuss, B., 2017. Robotic green asparagus selective harvesting. IEEE/ASME Transactions on Mechatronics 22, 2401–2410. doi:10.1109/TMECH.2017.2735861.

Liu, X., Wang, D., Li, Y., Guan, X., Qin, C., 2022. Detection of green asparagus using improved mask r-cnn for automatic harvesting. Sensors 22, 9270. doi:10.3390/s22239270.

Lund, W.J., 1985. Asparagus harvester. URL: https://patents.google.com/patent/US4512145A/en.

Lund, W.J., 2022. Asparagus severing assembly with heated valve unit, asparagus lifting apparatus with asparagus spear guides and interfering member, and asparagus harvester formed therewith. URL: https://patents.google.com/patent/US20220015294A1/en.

Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., Liu, Y., 2024. Rt-detrv22: Improved baseline with bag-of-freebies for real-time detection transformer. arXiv preprint arXiv:2407.17140 doi:10.48550/arXiv.2407.17140.

MARNE, M.S., WARBICK, J., TASNIM, S., YANG, S., 2021. Automated asparagus harvester: survey and comparative analysis. 5th International

Conference of the International Commission of Agricultural and Biosystems Engineering (CIGR). URL: https://library.csbe-scgab.ca/docs/meetings/2021/CSBE21908.pdf. accessed on November 10, 2024.

McInnes, L., Healy, J., Astels, S., 2017. hdbscan: Hierarchical density based clustering. Journal of Open Source Software 2, 205. doi:10.21105/joss.00205.

Mears, D., Moore, M., 1969. The potential of mechanical asparagus harvesters. TRANSACTIONS of the ASAE 12, 813–0815. doi:10.13031/2013.38962.

Mehta, S., 2023. Microsoft's azure kinect developer kit technology transfers to partner ecosystem. Tech Community Microsoft Blog. URL: https://techcommunity.microsoft.com/t5/mixed-reality-blog/microsoft-s-azure-kinect-developer-kit-technology-transfers-to/ba-p/3899122. accessed on November 10, 2024.

Miranda, J.C., Arnó, J., Gené-Mola, J., Lordan, J., Asín, L., Gregorio, E., 2023. Assessing automatic data processing algorithms for rgb-d cameras to predict fruit size and weight in apples. Computers and Electronics in Agriculture 214, 108302. doi:10.1016/j.compag.2023.108302.

Moore, M.J., Carpenter, H.E., 1968. Asparagus harvesting field tests, in: ASAE Paper No. NA-68-211.

Mu, S., Dai, N., Yuan, J., Liu, X., Xin, Z., Meng, X., 2024. S2cpl: A novel method of the harvest evaluation and subsoil 3d cutting-point location for selective harvesting of green asparagus. Computers and Electronics in Agriculture 225, 109316. doi:10.1016/j.compag.2024.109316.

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. doi:10.3390/agronomy11091780.

Peebles, M., Barnett, J., Duke, M., Lim, S.H., 2020. Robotic harvesting of asparagus using machine learning and time-of-flight imaging–overview of development and field trials, in: 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), pp. 1361–1366. doi:10.1109/CASE48305.2020.9217006.

Peebles, M., Lim, S., Duke, M., McGuinness, B., 2019. Investigation of optimal network architecture for asparagus spear detection in robotic harvesting. IFAC-PapersOnLine 52, 283–287. doi:10.1016/j.ifacol.2019.12.535.

Peebles, M., Lim, S.H., Duke, M., Au, C.K., 2018. Overview of sensor technologies used for 3d localization of asparagus spears for robotic harvesting. Applied Mechanics and Materials 884, 77–85. doi:10.4028/www.scientific.net/AMM.884.77.

Peebles, M., Lim, S.H., Duke, M., McGuinness, B., Au, C.K., 2024. Localization of asparagus spears using time-of-flight imaging for robotic harvesting. Industrial Robot: The International Journal of Robotics Research and Application doi:10.1108/IR-01-2024-0009.

Peebles, M.C.S., 2021. Time-of-flight perception pipeline for selective green asparagus harvesting: Theory and application. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand. URL: https://hdl.handle.net/10289/14605.

Read, C.M., 2009. Asparagus growth model. Master Thesis, Washington State University, Pullman, WA, USA. URL: https://hdl.handle.net/2376/100242.

Rigney, M., Brusewitz, G., Kranzler, G., 1992. Asparagus defect inspection with machine vision. Transactions of the ASAE 35, 1873–1878. doi:10.13031/2013.28809.

Sakai, H., Shiigi, T., Kondo, N., Ogawa, Y., Taguchi, N., 2013. Accurate position detecting during asparagus spear harvesting using a laser sensor. Engineering in Agriculture, Environment and Food 6, 105–110. doi:10.1016/S1881-8366(13)80019-5.

Suresh Kumar, M., Mohan, S., 2023. Selective fruit harvesting: Research, trends and developments towards fruit detection and localization–a review. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 237, 1405–1444. doi:10.1177/09544062221128443.

USDA-NASS, 2024. Vegetables 2023 summary. United States Department of Agriculture National Agricultural Statistics Service: Washington, DC, USA. URL: https://downloads.usda.library.cornell.edu/usda-esmis/files/02870v86p/qz20vd735/ht24z584t/vegean24.pdf. accessed on November 10, 2024.

Vasconez, J.P., Delpiano, J., Vougioukas, S., Cheein, F.A., 2020. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. Computers and Electronics in Agriculture 173, 105348. doi:10.1016/j.compag.2020.105348.

Vázquez-Arellano, M., Griepentrog, H.W., Reiser, D., Paraforos, D.S., 2016. 3-d imaging systems for agricultural applications–a review. Sensors 16, 618. doi:10.3390/s16050618.

Šlajpah, S., Munih, M., Mihelj, M., 2023. Mobile robot system for selective asparagus harvesting. Agronomy 13, 1766. doi:10.3390/agronomy13071766.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G., 2024a. Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458 doi:10.48550/arXiv.2405.14458.

Wang, C., Yeh, I., Liao, H., 2024b. Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616 doi:10.48550/arXiv.2402.13616.

Xu, J., Lu, Y., Deng, B., 2024. Openweedgui: An open-source graphical tool for weed imaging and yolo-based weed detection. Electronics 13, 1699. doi:10.3390/electronics13091699.

Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., Cortelazzo, G.M., 2016. Time-of-flight and structured light depth cameras: Technology and Applications. Springer. doi:10.1007/978-3-319-30973-6.

Zhang, P., Dai, N., Liu, X., Yuan, J., Xin, Z., 2024. A novel lightweight model hgca-yolo: Application to recognition of invisible spears for white asparagus robotic harvesting. Computers and Electronics in Agriculture 220, 108852. doi:10.1016/j.compag.2024.108852.

Zhang, P., Liu, X., Yuan, J., Liu, C., 2022. Yolo5-spear: A robust and real-time spear tips locator by improving image augmentation and lightweight network for selective harvesting robot of white asparagus. Biosystems Engineering 218, 43–61. doi:10.1016/j.biosystemseng.2022.04.006.

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., Chen, J., 2024. Detrs beat yolos on real-time object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition , 16965–16974doi:10.48550/arXiv.2304.08069.

Zhou, Q.Y., Park, J., Koltun, V., 2018. Open3d: A modern library for 3d data processing. doi:10.48550/arXiv.1801.09847.