Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413

Article

Comparing RGB-D Sensors for Close Range Outdoor Agricultural Phenotyping

Adar Vit [1], **Dr. Guy Shani** [2]

[1] Software and Information Systems Engineering, Ben Gurion University; adarv@post.bgu.ac.il
[2] Software and Information Systems Engineering, Ben Gurion University; shanigu@bgu.ac.il

Abstract: Phenotyping is the task of measuring plant attributes for analyzing the current state of the plant. In agriculture, phenotyping can be used to make decisions concerning the management of crops, such as the watering policy, or whether to spray for a certain pest. Currently, large scale phenotyping in fields is typically done using manual labor, which is a costly, low throughput process. Researchers often advocate the use of automated systems for phenotyping, relying on the use of sensors for making measurements. The recent rise of low cost, yet reasonably accurate, RGB-D sensors has opened the way for using these sensors in field phenotyping applications. In this paper, we investigate the applicability of 4 different RGB-D sensors for this task. We conduct an outdoor experiment, measuring plant attribute in various distances and light conditions. Our results show that modern RGB-D sensors, in particular, the Intel D435 sensor, provides a viable tool for close range phenotyping tasks in fields.

Keywords: RGB-D sensors; Empirical analysis; Sensors in agriculture; Phenotyping; Microsoft kinect; Intel D-435; Intel SR300; Orbbec astra s

1. Introduction

The constant increase of the world's population increases the needs for technological developments in agriculture industry. In order to meet the growing demand for food, the agriculture industry has to develop technological tools which will allow us to increase crop production [1]. The process of crop phenotyping, including the extraction of visual traits from plants, allows farmers to examine their crops and infer important properties concerning the crop status, such as insufficient irrigation, or developing diseases [2].

Consequently, there is an urgent need for the development of novel methods in phenomics for a non-destructive determination of diverse traits under field conditions [3]. Non-destructive, meaning that phenotypic data can be collected from the same organism over the course of a long experiment. They are also amenable to automation, making it feasible to study large sample sizes for increased statistical power. Image-based phenotyping approach aims to perform extraction phenomics based on obtained images data in non-destructive way [4]. Recently, many new technologies use computer vision methods for phenotyping [5].

Field and greenhouse phenotyping — measuring plant phenotypes in standard growing conditions, rather than in a controlled lab, is a difficult challenge [6]. Field conditions are notoriously heterogeneous and the inability to control environmental factors such as lighting, makes results difficult to interpret, however, results from controlled environments are far removed from the situation plants will experience in the field and, therefore, are difficult to extrapolate to the field [2]. Furthermore, given the slim profit margins in agriculture, farmers cannot invest in costly sensors for phenotyping.

Low-cost sensors can be used by the farmer, whether in a field or in a greenhouse, in order to allow for affordable phenotyping.

Indeed, many low cost visual sensors were developed in the last decade. For phenotyping, where one often needs to measure size, length, or width, RGB-D sensors are a viable alternative. RGB-D sensors provide, in addition to the three color channels (RGB – Red, Green, Blue), a depth channel, measuring the distance from the sensor to a point in the image. Using this information we can estimate, e.g., the length and width of a stem, or the size of a fruit.

There are several alternative technologies on which low cost RGB-D sensors are based. Optical techniques for image acquisition can be divided to passive and active. Passive methods use the reflection of natural light on a given target to measure its shape. Active methods enhance shape acquisition by using an external lighting source that provides additional information [7]. For example, time-of-flight (TOF) sensors measure depth by estimating the time delay from light emission to light detection. Structured-light sensors combine the projection of a light pattern with a standard 2D camera and measure depth via triangulation [8]. Active stereoscopy sensors looks for artificially projected features from multiple cameras to reconstruct a 3D shape, using triangulation and epipolar geometry theory [7]. Our work focuses on active sensors.

In this paper we evaluate 4 different low cost RGB-D sensors for agricultural phenotyping tasks. We compare the following sensors: MICROSOFT KINECT II, ORBBEC ASTRA S, INTEL SR300, and INTEL D435. We evaluate all sensors in identical outdoor settings, taking measurements of 6 corn plants, and 2 tomatoes of different colors and sizes. In addition, we have taken measurements of 2 plastic balls of different colors and sizes, as two reference objects of known diameter. Measurements were taken throughout the day in varying lighting conditions, and from varying distances, ranging from 20cm to 150cm.

We analyze the depth information received from the sensors to identify the fill rate — the portion of missing depth information over the objects of interest. The gathered images were used to train deep learning segmentation models for all object types (corn stalks, tomatoes, balls), which is an important task in phenotyping, and we report the quality of the resulting models. In addition, we have trained segmentation models using the depth information to estimate whether depth can enhance the quality of the models. We also use the depth information to estimate the known diameter of the plastic balls, and we compute the error of the various sensors.

Our results indicate that both the MICROSOFT KINECT II and the INTEL D435 provide much superior results to the INTEL SR300 and the ORBBEC ASTRA S in our settings. For many phenotyping applications, such as mounting sensors on drones [9,10], the lower weight and reduced energy demands of the INTEL D435 make it an attractive choice for agricultural outdoor applications.

2. Related Work

We now review related work. We begin with reviewing some research on the use of RGB-D sensors in agricultural applications, and then discuss research on comparing RGB-D sensors, both in general applications and specifically in agricultural settings.

2.1. Using RGB-D in Agriculture Applications

In recent years, the use of RGB-D sensors has been increasing due to the ability of these sensors to simultaneously capture depth and color images of the scene. Efficiently mapping the depth and color data, results in a colored point cloud in a 3-D spatial domain [11], which is very useful in many applications.

Chéné *et al.* [12] showed the potential and ability of RGB-D imaging systems for 3D measurements in the context of plant phenotyping, in what appears to be the first published application with plants. They showed that Kinect can resolve individual leaves, allowing automated measurement of leaf orientation in indoor environment.

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413

A wide variety of phenotyping tasks have been approached using RGB-D data. Xia *et al.* [13] presented an algorithm for leaf segmentation using RGB-D data. Using images of greenhouse paprika leaves, they demonstrate their ability to capture leaf measurements using an RGB-D camera. In [14] estimated the size of sweet onions in indoor conditions, and [15] used the RGB-D camera for size estimation of mango fruits on trees in field conditions. [16] showed the ability to measure the canopy structure of small plants in the field. Azzari *et al.* [17] used the Kinect sensor to characterize the vegetation structure. Jiang *et al.* [18] presented an algorithm for accurately quantifying cotton canopy size in field conditions. They showed that the multidimensional traits and multivariate traits were better yield predictors than traditional univariate traits, confirming the advantage of using 3D imaging modalities. [19] presented 3D visual detection method for detecting peduncles of sweet peppers in the field using short range depth camera Intel Real Sense F200.

Only a handful of studies use depth sensors for cereals phenotyping. Cereal are especially challenging, due to their relatively narrow stem, leaves, and head. Sodhi *et al.* [20] presented an automated method of mapping 2D images collected in an outdoor sorghum field and greenhouse to segmented 3D plant units that are of interest for phenotyping, using a multi-camera sensor.

2.2. Object Detection in Agriculture

Object detection is the task of identifying the positions of instances of semantic objects of a certain class in a given image using computer vision algorithms. Detection objects in an image is a crucial aspect in the development of agricultural applications. In order to harvest fruit or vegetable, navigate in the field, or spray selectively, the object location in an image has to be determined. This allows us to position robotic arms accordingly, to identify obstacles, or to estimate the object's characteristics such as ripeness or size [21].

Despite many years of research in agricultural oriented object detection, there are still many problems that hinder implementation of object detection in agricultural applications [22]. The highly variable and uncertain outdoor environment with changing illumination conditions, along with the complex plant structure and variable product shape and size make it hard to find a global solution to the detection of objects in the complex and unstructured agricultural environment [21].

2.3. Comparing RGB-D Sensors

RGB-D sensors were compared in several different studies, using varying metro-logical methods to evaluate the performance of the sensors under certain conditions and environments. For example, Sabattini *et al.* [23] and . Beltran and Basañez [24] compared the Kinect v1, which is based on structured-light technology, vs the PointGrey Bumblebee2 Stereo–camera for localizing a mobile robot. The performance of the two cameras was compared using a mobile robot and 2D landmarks. Their work demonstrate experimentally that stereo cameras have a smaller error in the determination of the 3D position of knows points in the image due to its better resolution.

Several studies [25–27], compared between two versions of Kinect: v1 and v2. Samir *et al.* [25] showed that the accuracy of v2 is slightly better than v1 with regards to the purpose of respiratory motion tracking, while Amon *et al.* [26] showed better performance for v2 in estimating the area of detection, as well as the rotation accuracy of two versions of the face tracking system of the Microsoft Kinect sensor. In [27], the comparison concentrated on depth image data. Their goal was to investigate the accuracy and precision of depth images of both devices. Accuracy is defined to be the difference or the offset of a measured depth value compared to a ground truth distance. Precision is defined as the repeatability of subsequent depth measurements under unchanged conditions. They investigate the influence of temperature, distance and object color on the captured depth images.

Diaz *et al.* [28] compares two generations of RGB-D sensors by evaluating the performances of Asus Xtion Pro, a structured light based camera, and Kinect v2. This evaluation considers the tasks of 3D reconstruction and object recognition. The quantitative comparison with respect to ground truth obtained using a metro-logical laser scanner, revealed that Kinect v2 provides less error in the mapping

between the RGB and depth frames, and the obtained depth values are more constant with distance variations.

Guidi *et al.* [29] compared between 5 low-cost 3D cameras with three different technologies by analyzing them in terms of systematic errors and random errors. They tested the cameras on a reference plain, made of a rectangular piece of float glass. Their tests have analyzed the range from 550mm to 1450mm, giving acceptable results with all the devices only between 550mm and 1150mm. Their results exhibit a global uncertainty similar for all the primesense-based devices, when the worst results are produced by the Realsense-based unit. They showed that the five low-cost 3D sensors they compared can certainly cater to gesture tracking and understanding.

The most profound and comprehensive was made in [7], they presented 20 3D camera commercially available that use varying technologies — structured light, time of flight, and active and passive stereoscopy. They focused on indoor metrological evaluations on the state of the art device in each technology: MICROSOFT KINECT II (time of flight), ORBBEC ASTRA S (structured light) and INTEL D435(active stereoscopy). They showed that the uncertainty in the depth measurement using a TOF camera scales linearly with the depth, hence providing reliable measurement at longer ranges. The ORBBEC ASTRA S and the Intel RS400TM, on the other hand, which are based on the triangulation principle, provide a depth measurement in which the uncertainty grows quadratically. Hence, their usage is preferred for short range applications.

Moreover, their work showed that the SR400TM generation of 3D camera provided by Intel proved to have outstanding performance when compared to other triangulation-based devices and for embedded applications, and that the RS400TM generation is a valuable device for shape acquisition. Although They used the same sensors as we do, we evaluate sensor performance in outdoor conditions and for agriculture phenotyping tasks.

Kazmi *et al.* [30] compared TOF cameras with stereo cameras for agricultural applications by evaluating close range depth imaging of leaves indoor and outdoor, under shadow and sunlight conditions, by varying the exposure of the sensors. Their evaluation metrics focused on analyzing the depth data by aggregating it across several frames using various statistical measures. Their work concludes that TOF cameras are sensitive to ambient light, and stereo vision is relatively more robust for outdoor lighting.

Wang *et al.* [15] estimate mango fruit size on trees in field conditions. Three low cost distance measurement technologies were compared for use in a fruit sizing system, to be mounted on a moving platform equipped with LED illumination for night imaging of mango orchards. For estimation of camera-to-fruit distance, low-cost examples of three distance measurement technologies were compared under lab conditions (fluorescent lighting): Zed (stereo vision camera), Leica (TOF laser) and the MICROSOFT KINECT II (TOF and RGB-D camera). The distance measurement was taken using three materials, varying in their level of diffuse reflectance, which placed placed at 14 positions ranging from 0.56 to 5.4 m. In addition, the ceramic tile for Kinect distance measurement was repeated outdoors at times from early afternoon to after sunset. The Zed stereo depth imaging technique was inferior to other technologies. In direct sunlight with a ceramic target, The Kinect failed to measure distance over 3.5m.

3. Materials and Methods

We now report the experiment that was used to collect the data. In the experiment, 4 different RGB-D sensors were used in an outdoor scenario to take measurements of young corn plants and tomatoes.

Figure 1. The setup of the 4 sensors used in the experiment. Kinect II at the bottom, the Astra Pro on the left, the SR300 on the right, and the gray D435 on top of the SR300.

3.1. RGB-D Sensors

We compare the following RGB-D sensors: ORBBEC ASTRA S[1], MICROSOFT KINECT II[2], INTEL SR300[3], and INTEL D435[4]. Table 1 shows the properties of all cameras. We now briefly review the sensors.

3.1.1. Astra S

The Astra S sensor is manufactured by Orbbec company in China. The camera is based on structured-light technology and it is designed for short range measurements. The structured-light technology uses a single camera with a structured pattern projected on the scene. An infra red (IR) projector projects a codified pattern embedding sufficient structure to provide unique correspondence. The direction of the structured pattern is known a priori, allowing triangulation based on the pattern [7]. The device contains an RGB sensor, IR sensors and a coded pattern projector. In addition, the device includes 2 microphones and an advanced eye protector.

3.1.2. Microsoft Kinect II

The MICROSOFT KINECT II sensor for Windows was released in 2014, along with a supporting SDK, allowing human body and face tracking. The device is based on Time-of-Flight (TOF) technology, estimating distance based on the known speed of light. Such sensors measure the time-of-flight of a light signal between the camera and the subject for each point of the image. The device requires a powerful illumination system with relatively high energy consumption. The device contains a full HD RGB camera which is registered with an IR camera. The device also includes IR emitters and a microphone.

3.1.3. Intel©RealSense™

Intel RealSense™ provides an open platform for developers to incorporate Intel's perceptual devices in their applications. The LibRealSense™ cross-platform API provides several tools for managing the sensor's streams, as well as advanced functions for background removal, hand tracking, fine face recognition and 3D scanning.

---

1 https://orbbec3d.com/product-astra/
2 https://developer.microsoft.com/en-us/windows/kinect
3 https://software.intel.com/en-us/realsense/sr300
4 https://realsense.intel.com/stereo/

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0664.v1

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413

3.1.4. Intel©SR300

Intel©SR300 was released in 2016 and it is the second generation of front-facing Intel©RealSense™ cameras. Similar to the Astra sensor, The device is based on structured light technology and it is also designed for short range. The SR300 is a subassembly camera product that implements a short range (SR), coded light, 3D imaging system. Along with an infrared laser projector, the subassembly includes a Fast VGA infrared camera and a 2-M pixel RGB color camera with an integrated image signal processor [31]. The cameras are factory-calibrated and the intrinsic and extrinsic parameters of the sensors are stored on board, easily accessible via the librealsense APIs.

3.1.5. Intel D435

INTEL D435is part of Intel's D400™ series, featuring the D435™ and the D415™. The Intel D435 depth camera is based on infrared active stereoscopy technology, with a global shutter sensor. The depth is estimated in hardware through an imaging ASIC that processes the infrared stream together with the RGB stream. The device performs frame correlation with a census cost function to identify homologous points and reconstructs the disparity. As it is based on active stereoscopy, an infrared dot-pattern projector adds textures to the scene, to cope with low-texture environments, where the D435 has random focused dot pattern. Both structured light and active stereoscopy are based on the same triangulation principle [7].

In our experiment we tested the D435 with 3 modes: 1280×720 pixel resolution, 848×480 pixel resolution and 640×480 pixel resolution.

Table 1. Properties of the RGB-D sensors used in our experiment.

	Sensor			
Parameter	**ORBBEC ASTRA S**	**MICROSOFT KINECT II**	**INTEL SR300**	**INTEL D435**
Range	0.4m-2m	0.5m-4.5m	0.3m-2m	0.2m -10m
RGB FOV	60°(H) × 49.5°(V) × 73°(D)	70.6°(H) × 60°(V)	41.5°(H) × 68°(V) × 75.2°(D)	69.4°(H) × 42.5°(V)) × 77°(D)
Depth FOV	60°(H) × 49.5°(V) × 73°(D)	70.6°(H) × 60°(V)	55°(H) × 71.55°(V) × 88°(D)	91.2°(H) × 65.5°(V)) × 100.6°(D)
Frame rate	30 fps	30 fps	30, 60 fps	30, 60, 90 fps
RGB resolution	640×480 pixel	1920 × 1080 pixel	1920 × 1080 pixel	1920 × 1080 pixel 1280×720 pixel 848×480 pixel 640×480 pixel
Depth resolution	640×480 pixel	512 × 424 pixel	640 × 480 pixel	1280 × 720 pixel 848×480 pixel 640×480 pixel
Weight	300 gr	966gr	300gr	100gr
Size	165mm × 30mm × 40mm	255mm × 66mm × 67mm	110mm × 12.6mm × 4.1mm	90mm × 25mm × 25mm
Power supply	USB 2.0	power supply + USB 3.0	USB 3.0	USB 3.0
Power consumption	< 2.4 W	∼ 15 W	650-1800 mW	618 -1978 mW
Operating system	Android, Linux Windows 7/8/10	Linux Windows 8/10	Linux Windows 8/10	Linux Window 8/10
SDK	Astra SDK OpenNI2 3rd party S	Kinect V2 SDK libfreenect2	Intel©RealSense™ SDK librealsense sdk [6]	Intel©RealSense™ SDK librealsense SDK[5] hand and face tracking

3.2. Measured Objects

To test the performance of the RGB-D sensors, we took measurements of 6 different young corn plants, organized in two rows of 3, to simulate the foreground-background setting in fields. The plants were 5 weeks old, and approximately 50 cm tall, and were planted in black plastic pots, allowing us to

Figure 2. The experimental setup as captured by the Kinect sensor from 1 meter. Six corn plants in black plastic pots arranged in two rows. Added tomatoes — a red tomato, an orange cherry tomato. Two plastic balls for size measurements — a larger green ball and a smaller yellow ball.

easily move them to different positions. The plants have not yet developed flowers or corn ears and husk.

In addition, we added two different types of tomatoes, allowing us to evaluate the usability of the captured images for object identification other than corn stems. We used one regular red tomato, and one oval orange cherry tomato, positioning them near the corn stalks.

In addition, for reliable measurements of object width, we added two plastic balls: a green ball with a diameter of 50 mm, and a smaller yellow ball with a diameter of 7.8 mm. The ball will be later used to evaluate the ability of the various RGB-D sensors for the important task of computing the width and length of an object of interest, such as stems, fruits, and leaves.

Figure 2 shows the setting of the corn plants, tomatoes, and plastic balls.

3.3. Procedure

In order to evaluate the performance of the RGB-D sensors for close range at various lighting conditions, we conducted an experiment in an outdoor environment. we assembled an imaging platform which employed a number of sensors: ORBBEC ASTRA S, MICROSOFT KINECT II, INTEL SR300and INTEL D435. The ORBBEC ASTRA Sand the INTEL D435sensors were positioned next to each other, directly above the MICROSOFT KINECT II. The INTEL D435sensor was positioned above the ORBBEC ASTRA S. The sensors assembly was fixed during the entire experiment, and the sensors were not moved. We used three different computers, allowing us to reduce the time between measurements of different sensors. For the Kinect v2 and the Astra s we used the official SDK for capturing RGB-D images, and the librealsense SDK was used for the Intel cameras.

As our goal is evaluation for short distances, our target objects were placed at 7 positions ranging from 0.2 to 1.5m. After measurement were taken, the pots were moved to the next position. The images acquisition process was repeated in twelve cycles, at various lighting conditions, from sunrise

to sunset. At each cycle we measured the luminous flux using the ambient light sensor of a Galaxy S8. The lux values that the sensor provided were categorized into 4 ranges [7]:

1. Sunrise light: up to 1000 lux.
2. Overcast lighting: from 1000 to 10000 lux.
3. Full daylight (vut not direct sun): from 10000 to 32000 lux.
4. Direct sunlight: 32000 lux and above.

4. Results

The main goal of our study is to evaluate the usability of the various sensors for outdoor close range phenotyping. Below, we provide the results over the conducted experiment. We report the amount of the depth information that was captured, analyzing the fill rate of the sensors in varying conditions. We then report the quality of the RGB and the depth information that was captured for the important task of object identification. We then analyze the usability of the depth information for estimating object size, which is also often useful in phenotyping applications.

4.1. Fill rate

It is often the case that some pixels in an depth image contain no depth information (typically marked as 0 depth), or contain incorrect depth measurements, such as distances which are much closer, or much farther, than the actual distance to the object of reference. The fill rate of a sensor is the portion of pixels that contain valid measurements within a region of interest (ROI). The fill rate is critical for tasks such as the segmentation of objects, or the measurement of width and length.

As we are interested in agricultural applications, our ROIs are the corn stems, the tomatoes, and the plastic balls. We calculate the fill rate of depth values in each ROI for each distance. We consider a pixel measurement to be valid if it has a non-zero value, and it is within 3 standard deviations from the mean of non-zero depth values in the ROI.

Figure 3 shows the fill rate over all objects of interest by the distance to the objects. The ORBBEC ASTRA S sensor produces the worst results here, with a fill rate of about 10% from 40cm to 100cm. The INTEL SR300 depth sensor is designed only for short range measurements. As such, it provides about 50% fill rate at 20cm, 10% at 40cm, and almost no depth information above that range. The MICROSOFT KINECT II depth sensor is designed to operate in a range above 50cm. Indeed, from 50cm and on, the MICROSOFT KINECT II sensor produces very good results, with about 90% fill rate. Finally, the INTEL D435 sensor, in all possible resolutions, provides the best overall performance, with the highest fill rate in short ranges (20cm to 40cm), and a comparable rate to MICROSOFT KINECT II in the range of 60cm to 150cm. This is impressive, especially given the difference in power consumption between the two sensors.

We now analyze the sensitivity of the depth sensors to the lighting conditions. This is especially important in the uncontrolled field conditions that we are interested in, where lighting conditions may vary considerably. For this analysis we consider only distance range where the sensors operate well — above 40cm for MICROSOFT KINECT II, below 40cm for INTEL SR300, 40cm to 100cm for ORBBEC ASTRA S, and all ranges for INTEL D435.

Figure 4a shows the fill rate by the time of day, and Figure 4 shows the fill rate by light intensity. As can be seen, the ORBBEC ASTRA S sensor operates best in the lowest lighting intensity measured, while the INTEL SR300 sensor operate best in the medium lighting conditions, reaching a fill rate comparable to the best sensors at one specific time of day — early morning with medium lighting intensity. Both the MICROSOFT KINECT II and the INTEL D435 sensors showed little sensitivity to the lighting conditions, with a slight decrease in performance in the highest lighting conditions measured, around mid-day.

[7] http://stjarnhimlen.se/comp/radfaq.html

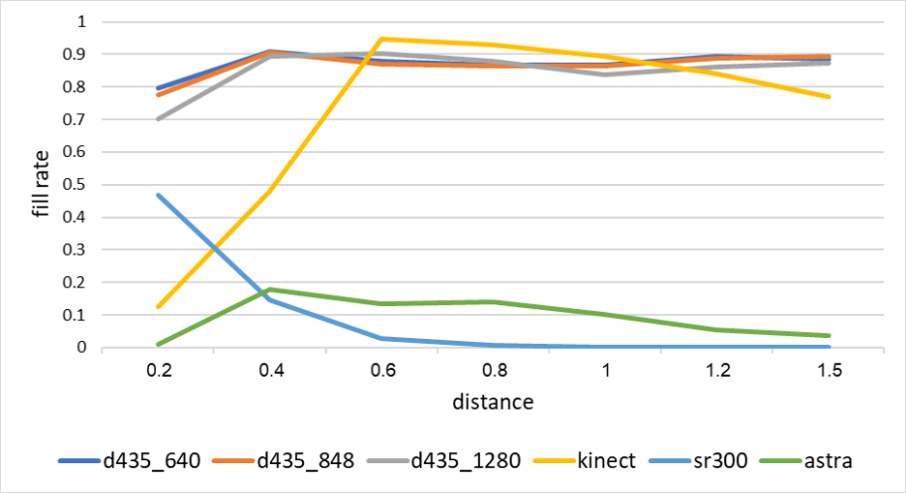


Figure 3. Mean fill rate over all objects of interest for each camera by distance to the objects.

There are several factors that can affect the fill rate percentage such as an object's shape and the reflection of light over the object. Figure 5 shows the fill rate for each object type, focusing on the two best sensors — MICROSOFT KINECT II and INTEL D435. The sensors vary in their behavior — the INTEL D435 captures the ball best, has medium fill rate over the tomatoes, and the lowest performance on the stems, while the MICROSOFT KINECT II has the best performance over the stems, and worse performance on both the tomatoes and the ball.

When considering the time of day, we see that the INTEL D435 is unaffected by the light intensity in different hours for the ball, but suffers a reduction in quality for stem fill rate. The MICROSOFT KINECT II suffers a considerable reduction in quality during the hours when the sun light is strongest, but this reduction is less noticeable for the stems.

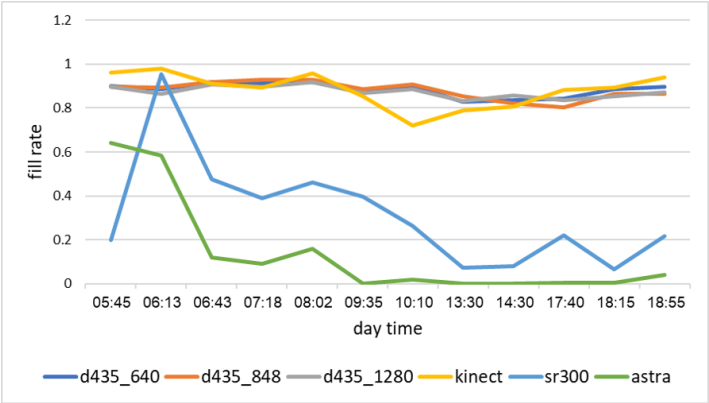
4.2. Object Detection using Deep-Learning

In order to compare the quality of the data produced by the sensors for the task of object detection, we use here a state-of-the-art algorithm, the Mask R-CNN model [32]. Our goal was to detect 3 different classes: the larger green ball, the tomato fruit, and the corn stem. In all the experiments below, we split the images by the time of day to train, validation, and test. There were 12 different time points where an image capture cycle for all distances has began. Images from 3 cycles were used for test, images from one cycle was used for validation, and images from the other 8 cycles were used for training the models. As we are interested in comparing the images on identical algorithmic settings, we did not perform any hyper-parameter tuning, training all models with the same algorithmic configuration. Each model was trained for 60 epochs, where on the last epoch contained also the validation set. The train set contained 126 images, the validation set 14 images and the test set 28 images.

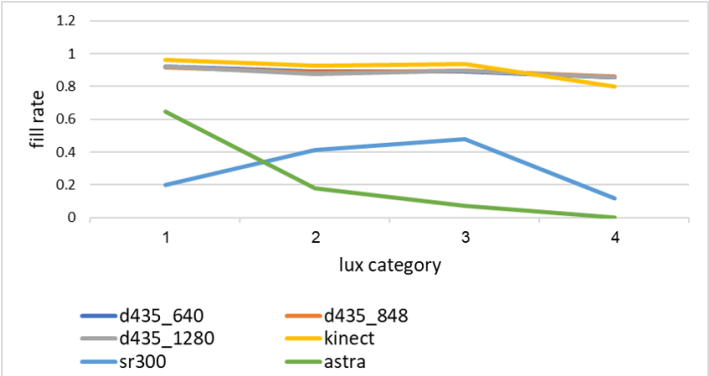
First, we examine the performance of all RGB sensors in the detection task by training a model for each camera, using only the obtained RGB images. Table 2 presents the mAP (mean average precision) results for each camera's model. As the table shows, the INTEL D435 RGB images produced the best results in the lowest resolution. The ORBBEC ASTRA S RGB data produced the lowest quality object identification model.

We also experimented with using the depth information produced by the sensors to augment the RGB data in object identification [33]. As only the MICROSOFT KINECT II and the INTEL D435 sensors produced a reasonable fill rate for most ranges, we limit this experiment only to data gathered using these two cameras. We used only images captured within the camera's working range, above 60cm for MICROSOFT KINECT II and above 40cm for the INTEL D435. Table 3 shows the image split into train, validation and test sets for each sensor.

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413



(a) Fill rate by daytime



(b) Fill rate by light intensity

Figure 4. Fill rate for all depth sensors, considering only objects within the camera range.

Table 2. mAP results for the object detection task using only RGB data.

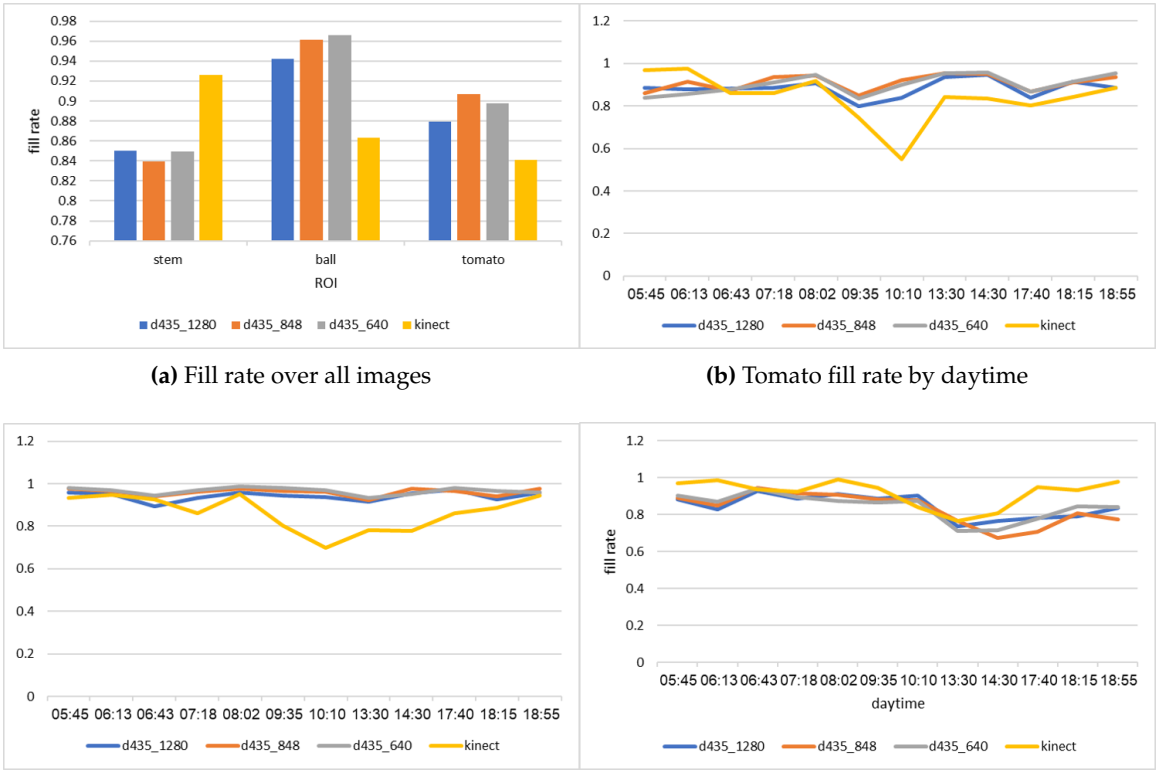
Sensor	mAP	Tomato	Ball	Stem
Intel D435: 640	0.955	1	0.99	0.87
Intel D435: 1280	0.906	1	1	0.72
Intel sr300	0.925	1	0.95	0.83
Intel D435: 848	0.912	1	1	0.74
Microsoft Kinect II	0.914	1	1	0.75
Orbbec Astra s	0.826	0.96	0.88	0.64

We experimented with two methods for incorporating the depth information. First, we replaced the blue channel in the RGB image with the depth information. We hence trade some color information for the depth information. As can be seen in Table 4, this method helped in only a two cases — the Intel D435 sensor with the highest resolution and the Microsoft Kinect II. For other Intel D435 resolutions this method only reduced the performance. This may be due to insufficient training data, as the original Mask R-CNN network that we used was trained with standard RGB data, and may be incompatible with this type of data.

We next replaced the image background with black pixels. As we know the distance d between the camera to the object of interest, we only maintain the color of pixels within 25 cm of d. To avoid removing pixels within the object of interest with missing or incorrect depth information, a pixel that had a neighbor within a 10 pixel radius that was in the 25 cm range, was not modified. The color of all other pixels was changed to black. As can be seen in Table 4 this method improved object identification in all cases. Figure 6 provides an example of object identification using the three methods. As we can see, the removed background prevented the identification of a background object.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0664.v1

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413



(a) Fill rate over all images

(b) Tomato fill rate by daytime

(c) Ball fill rate by daytime

(d) Stem fill rate by daytime

Figure 5. Comparing fill rate for different objects

Table 3. Train, validation, and test splits for depth integration in object detection.

Set	Microsoft Kinect II	Intel D435
Train	86	112
Validation	8	10
Test	18	22

Table 4. mAP results for object identification incorporating depth information for all object types. RGB denotes using only the RGB data, BKGD denotes background removal prior to identificaiton, and Blue denotes replacing the blue channel with depth information.

	All			Tomato			Ball			Stem		
Sensor	RGB	BKGD	Blue	RGB	BKGD	Blue	RGB	BKGD	Blue	RGB	BKGD	Blue
Kinect	0.866	0.924	0.9	1	1	1	0.9	0.96	0.9	0.7	0.81	0.81
D435:1280	0.912	0.932	0.938	0.97	0.97	0.98	0.92	0.95	0.92	0.85	0.87	0.92
D435:848	0.962	0.971	0.936	1	1	1	1	1	1	0.89	0.91	0.81
D435:640	0.976	0.981	0.953	1	1	1	1	1	1	0.93	0.94	0.86

4.3. Object Size Estimation

Measuring the size of a fruit [15], or the width of a leaf or a stem [34] can be an important phenotype which can indicate the plant condition. We therefore examined the capabilities of the sensors to approximate the size of two different objects of reference: the two plastic balls that were used. We choose the balls rather than the actual stems and fruits to avoid confusion from, e.g., varying width at different points along the stem. The larger ball had a diameter of 50mm, which is similar to the size of a tomato, while the smaller ball had a diameter of 7.8mm, similar to the width of a corn stem.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0664.v1

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413

(a) RGB **(b)** BKGD **(c)** Blue

Figure 6. INTEL D435:848 detection examples for each method. The blue bounding box denotes the ground truth, the green bounding box denotes a true positive prediction, and the red bounding box denotes a false positive prediction.

Given the distance of an object from the sensor, one can approximate its size or length. Given a twp pixels in the image, and the depth data, one can compute their 3D coordinates [7]:

$$X = \frac{D_{x,y} \cdot (c_x - x)}{f_x} \quad (1)$$

$$Y = \frac{D_{x,y} \cdot (c_y - y)}{f_y} \quad (2)$$

$$Z = D_{x,y} \quad (3)$$

where $D_{x,y}$ is the distance to the pixel at coordinates x, y, c_x and c_y are the principal points and f_x and f_y are the focal lengths expressed in pixel units. We used $D_{x,y}$ as the average distance of two selected pixels.

Given the 3D coordinates for each pixel, measuring the distance between two pixels is simple, using Euclidean distance in 3D, denoted $d(p_1, p_2)$. Given a series of images, let p_1^i and p_2^i be two pixels on the opposite sides on the perimeter of the ball in image i. We can now compute the root of the mean square error (RMSE) of the estimation:

$$RMSE = \sqrt{\frac{\sum_{i=1..N}(d(p_1^i, p_2^i) - \Phi)^2}{N}} \quad (4)$$

where Φ is the true ball diameter. In some cases it is interesting to compute relative error with respect to the size of the object. This is because an error of, e.g., 1cm, may be problematic when estimating a distance of 10cm, but may be considered negligible when estimating a distance of 1 meter. We hence compute the mean relative average error (MRAE):

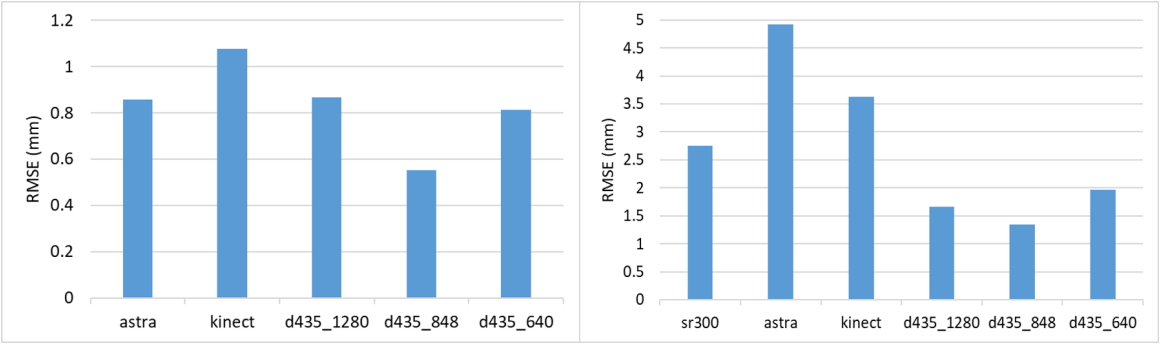
$$MRAE = \frac{\sum_{i=1..N} \frac{(d(p_1^i, p_2^i) - \Phi)^2}{\Phi}}{N} \quad (5)$$

For estimating the ball diameter using the various sensors, we manually selected two pixels, positioned on two opposite sides of the ball perimeter. We then use Equations 1 to 3 to compute the 3D coordinates of the two pixels, and finally computed the Euclidean distance between the points.

Figure 7 shows the RMSE results for estimating the diameter of the two balls. As can be seen, all INTEL D435 resolutions performed better than all other sensors. For both balls, the 848 resolution provided the best depth information, and hence, the best diameter estimation. The ORBBEC ASTRA S sensor had the worst estimation with the highest error on the larger ball. This is because the ORBBEC ASTRA S sensor did not capture well the distance to points on the perimeter of the larger ball from the perspective of the camera, possibly due to reflection angles. For the smaller ball, the ORBBEC ASTRA S

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0664.v1

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413



(a) Small ball **(b)** Large ball

Figure 7. RMSE for diameter estimation of the two balls.

sensor provided much better estimations. The INTEL SR300 sensor preformed well on the larger ball, but it is limited only to close range, and was able to provide estimations only within a 40cm distance, and was unable to capture the smaller ball at any range.

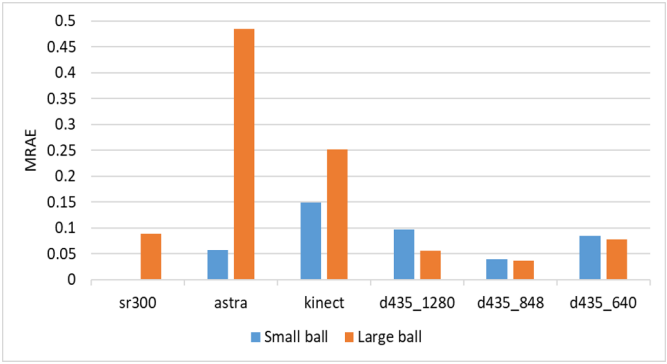


Figure 8. MRAE for diameter estimation of the two balls.

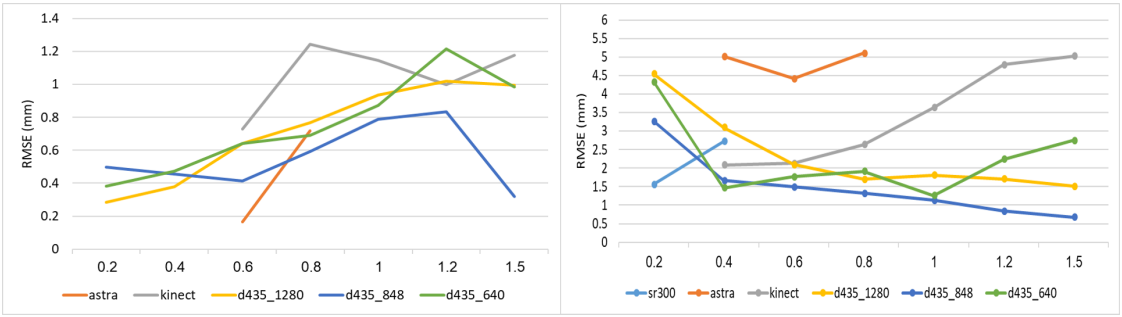
Figure 8 shows the relative error (MRAE) with respect to the ball diameter. We can see here that the INTEL D435 sensor provided the best relative error, below 5% of the object diameter for the 848 resolution, and below 10% for the other resolutions.

Depth estimation often depends on the distance from the sensor to the object. Figure 9 shows the RMSE for varying object distances. The results are different for the various sensors. For the larger ball, the MICROSOFT KINECT II sensor error grows with the distance to the target object, but the INTEL D435 error is reduced with the distance, and for the best resolution, 848, it is at its lowest at the 1.5 meter range. The INTEL SR300 works best at the 20cm range, where it provides the best results, and reasonably well for the 40cm range. The ORBBEC ASTRA S sensor works poorly on all ranges here. For the smaller ball the results are different, and the INTEL D435 errors are not well correlated with the distance for the 848 resolution, while the error grows with the distance for the 1280 and the 640 resolution. For the smaller ball, the ORBBEC ASTRA S sensor provides the best results at the 60cm range, but much worse results for the 80cm range.

Another factor that may affect distance estimation is the light intensity. Figure 10 shows the distance estimations in various lux categories. For the larger ball, the INTEL D435 sensor is almost unaffected by light intensity. For the smaller ball, the INTEL D435 estimation is worse as the light intensity grows. The ORBBEC ASTRA S sensor preforms the best at the lowest lux for the smaller ball, and is competitive at the lowest lux for the larger ball as well, but its performance is reduced substantially as the light intensity grows, failing to capture any meaningful depth information in the

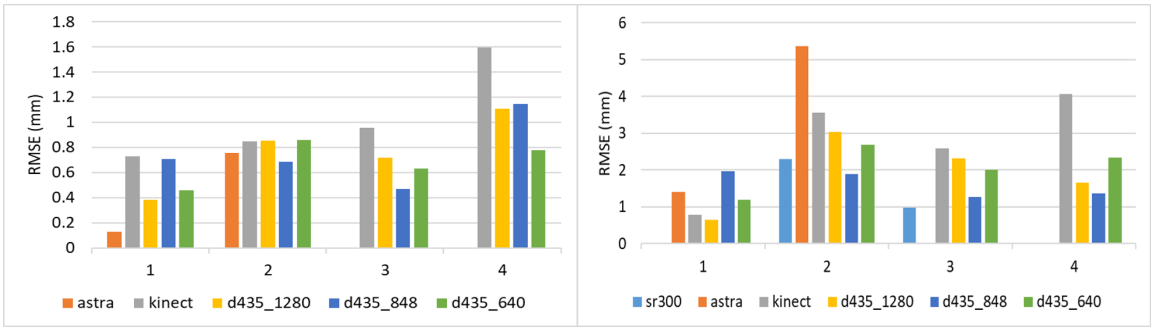
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0664.v1

Peer-reviewed version available at Sensors 2018, 18, 4413; doi:10.3390/s18124413



(a) Small ball **(b)** Large ball

Figure 9. RMSE for diameter estimation of the two balls from varying distances.



(a) Small ball **(b)** Large ball

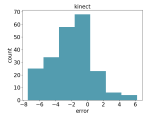
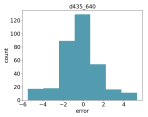
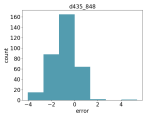
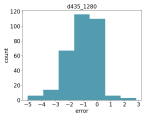
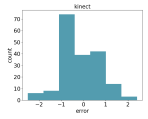
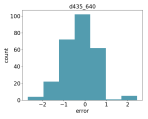
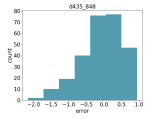
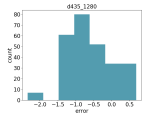
Figure 10. RMSE for diameter estimation of the two balls in varying light intensity categories.

higher lux categories. The MICROSOFT KINECT II sensor provides competitive results at lower light intensities, but its performance degrades at higher light intensities.

As we have seen above, the fill rate of the INTEL SR300 and ORBBEC ASTRA S are low. Hence, the amount of images containing sufficient depth information obtained by these two sensors (objects with a fill rate over 50%) is low compared with the MICROSOFT KINECT II and the INTEL D435 sensors. Moreover, ORBBEC ASTRA S and INTEL D435 are limited to shorter ranges. We will therefore separate analysis of the sensors.

Finally, in many cases of agricultural applications, one is not necessarily interested in estimating the size of a specific object, but rather in the average size of objects of a given type in an area. For example, one may wish to compute the average tomato size in a greenhouse, or the average stem width in a corn field. In these cases, knowing the error distribution can help to compensate for errors

Table 5. Error distribution

Ball	Kinect	D435:640	D435:848	D435:1280
Large				
Small				

in computing a better average estimation. Table 5 shows the distribution of errors for the INTEL D435 and MICROSOFT KINECT II sensors. As can be seen, the errors are in general Gaussian, and are often biased towards either positive or negative errors. In a future study, we will measure a large number of objects of a given type, and analyze different methods for taking the measurement noise into account.

5. Discussion

All the analysis provided in the previous section clearly shows that the new INTEL D435 sensor can be very useful in field agricultural applications. The INTEL D435 produces RGB data of sufficient quality for object identification using image-based deep learning methods, which is an important task in agriculture. The depth information produced by the sensor also has the highest quality of the sensors that we experiment with. First, the INTEL D435 is able to produce a competitive fill rate over all the ranges in our experiments — from 20cm to 1.5 meters. Furthermore, when using the depth measurements for object size estimation, which is also an important task in agriculture, the INTEL D435 sensor provided the best results.

The INTEL D435 sensor supports 3 different resolutions. In our experiments, the 848 resolution provided the best depth information, resulting in lower errors in object size estimation. The RGB information of the lower 640 resolution provided better object identification capabilities. The better performance for object identification may be attributed to the properties of the deep learning algorithm, rather than the more accurate RGB data.

The MICROSOFT KINECT II sensor also provides a competitive fill rate at all ranges, but the depth measurement quality is lower at the larger distances, and also at higher light intensities. In addition, the MICROSOFT KINECT II has substantially higher energy requirements than the other sensors, which may pose a problem in applications that require a low weight sensor, such as in small drone applications.

The INTEL SR300 sensor provides competitive results at lower ranges, especially for measuring larger objects sizes, but it is limited only for close range applications.

Finally, the ORBBEC ASTRA S sensor did not preform well in our experiments, and may not be the best choice for outdoor applications.

6. Conclusion and Future Research

In this research we experiment with 4 RGB-D sensors — MICROSOFT KINECT II, ORBBEC ASTRA S INTEL SR300, and INTEL D435 with 3 different resolutions — in outdoor agricultural-oriented tasks. We compute the fill rate of the sensors over various agricultural objects, and estimate the quality of the depth information for measuring object size. We also use both the RGB and the depth data captured by the sensors for object identification using deep learning methods.

Our experiments clearly show that the new INTEL D435 sensor provides the best results in our settings, and is hence a viable alternative for outdoor agricultural applications. The relatively high quality depth information, together with low energy requirements, and low weight, make the INTEL D435 useful for many field applications, such as drone-based phenotyping tasks.

In the future we intend to use the INTEL D435 in field phenotyping tasks, such as measuring the average diameter of stems in wheat, corn, and tomatoes in the field, and also to estimate the average fruit size in various crops. We also intend to mount the INTEL D435 on a drone to test its applicability for drone-based phenotyping applications.

Acknowledgements

This research is supported by the Israel Innovation Authority through the Phenomics MAGNET Consortium.

1. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences* **2011**, *108*, 20260–20264.
2. Araus, J.L.; Cairns, J.E. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in plant science* **2014**, *19*, 52–61.
3. Mutka, A.M.; Bart, R.S. Image-based phenotyping of plant disease symptoms. *Frontiers in plant science* **2015**, *5*, 734.
4. Busemeyer, L.; Mentrup, D.; Möller, K.; Wunder, E.; Alheit, K.; Hahn, V.; Maurer, H.P.; Reif, J.C.; Würschum, T.; Müller, J.; others. BreedVision—A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors* **2013**, *13*, 2830–2847.
5. Fahlgren, N.; Gehan, M.A.; Baxter, I. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current opinion in plant biology* **2015**, *24*, 93–99.
6. An, N.; Palmer, C.M.; Baker, R.L.; Markelz, R.C.; Ta, J.; Covington, M.F.; Maloof, J.N.; Welch, S.M.; Weinig, C. Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Computers and Electronics in Agriculture* **2016**, *127*, 376–394.
7. Giancola, S.; Valenti, M.; Sala, R. A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies, 2018.
8. Sarbolandi, H.; Lefloch, D.; Kolb, A. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer vision and image understanding* **2015**, *139*, 1–20.
9. Stehr, N.J. Drones: The newest technology for precision agriculture. *Natural Sciences Education* **2015**, *44*, 89–91.
10. Pederi, Y.; Cheporniuk, H. Unmanned Aerial Vehicles and new technological methods of monitoring and crop protection in precision agriculture. Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), 2015 IEEE International Conference. IEEE, 2015, pp. 298–301.
11. Han, J.; Shao, L.; Xu, D.; Shotton, J. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics* **2013**, *43*, 1318–1334.
12. Chéné, Y.; Rousseau, D.; Lucidarme, P.; Bertheloot, J.; Caffier, V.; Morel, P.; Belin, É.; Chapeau-Blondeau, F. On the use of depth camera for 3D phenotyping of entire plants. *Computers and Electronics in Agriculture* **2012**, *82*, 122–127.
13. Xia, C.; Wang, L.; Chung, B.K.; Lee, J.M. In situ 3D segmentation of individual plant leaves using a RGB-D camera for agricultural automation. *Sensors* **2015**, *15*, 20463–20479.
14. Wang, W.; Li, C. Size estimation of sweet onions using consumer-grade RGB-depth sensor. *Journal of Food Engineering* **2014**, *142*, 153–162.
15. Wang, Z.; Walsh, K.B.; Verma, B. On-Tree Mango Fruit Size Estimation Using RGB-D Images. *Sensors* **2017**, *17*, 2738.
16. Hui, F.; Zhu, J.; Hu, P.; Meng, L.; Zhu, B.; Guo, Y.; Li, B.; Ma, Y. Image-based dynamic quantification and high-accuracy 3D evaluation of canopy structure of plant populations. *Annals of botany* **2018**, *121*, 1079–1088.
17. Azzari, G.; Goulden, M.L.; Rusu, R.B. Rapid characterization of vegetation structure with a Microsoft Kinect sensor. *Sensors* **2013**, *13*, 2384–2398.
18. Jiang, Y.; Li, C.; Paterson, A.H.; Sun, S.; Xu, R.; Robertson, J. Quantitative Analysis of Cotton Canopy Size in Field Conditions Using a Consumer-Grade RGB-D Camera. *Frontiers in plant science* **2018**, *8*, 2233.
19. Sa, I.; Lehnert, C.; English, A.; McCool, C.; Dayoub, F.; Upcroft, B.; Perez, T. Peduncle detection of sweet pepper for autonomous crop harvesting—Combined Color and 3-D Information. *IEEE Robotics and Automation Letters* **2017**, *2*, 765–772.
20. Sodhi, P.; Vijayarangan, S.; Wettergreen, D. In-field segmentation and identification of plant structures using 3D imaging. Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on. IEEE, 2017, pp. 5180–5187.
21. Kapach, K.; Barnea, E.; Mairon, R.; Edan, Y.; Ben-Shahar, O. Computer vision for fruit harvesting robots–state of the art and challenges ahead. *International Journal of Computational Vision and Robotics* **2012**, *3*, 4–34.
22. Gongal, A.; Amatya, S.; Karkee, M.; Zhang, Q.; Lewis, K. Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture* **2015**, *116*, 8–19.

23. Sabattini, L.; Levratti, A.; Venturi, F.; Amplo, E.; Fantuzzi, C.; Secchi, C. Experimental comparison of 3D vision sensors for mobile robot localization for industrial application: Stereo-camera and RGB-D sensor. Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on. IEEE, 2012, pp. 823–828.
24. Beltran, D.; Basañez, L. A Comparison between Active and Passive 3D Vision Sensors: BumblebeeXB3 and Microsoft Kinect. ROBOT2013: First Iberian Robotics Conference. Springer, 2014, pp. 725–734.
25. Samir, M.; Golkar, E.; Rahni, A.A.A. Comparison between the Kinect™ V1 and Kinect™ V2 for respiratory motion tracking. Signal and Image Processing Applications (ICSIPA), 2015 IEEE International Conference on. IEEE, 2015, pp. 150–155.
26. Amon, C.; Fuhrmann, F.; Graf, F. Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2. Proceedings of the 6th Congress of the Alps Adria Acoustics Association, 2014, pp. 16–17.
27. Wasenmüller, O.; Stricker, D. Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. Asian Conference on Computer Vision. Springer, 2016, pp. 34–45.
28. Diaz, M.G.; Tombari, F.; Rodriguez-Gonzalvez, P.; Gonzalez-Aguilera, D. Analysis and evaluation between the first and the second generation of RGB-D sensors. *IEEE Sensors journal* **2015**, *15*, 6507–6516.
29. Guidi, G.; Gonizzi, S.; Micoli, L.; others. 3D capturing performances of low-cost range sensors for mass-market applications. *INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES* **2016**, pp. 33–40.
30. Kazmi, W.; Foix, S.; Alenyà, G.; Andersen, H.J. Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison. *ISPRS journal of photogrammetry and remote sensing* **2014**, *88*, 128–146.
31. Carfagni, M.; Furferi, R.; Governi, L.; Servi, M.; Uccheddu, F.; Volpe, Y. On the performance of the Intel SR300 depth camera: metrological and critical characterization. *IEEE Sensors Journal* **2017**, *17*, 4508–4519.
32. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017, pp. 2980–2988.
33. Gupta, S.; Girshick, R.; Arbeláez, P.; Malik, J. Learning rich features from RGB-D images for object detection and segmentation. European Conference on Computer Vision. Springer, 2014, pp. 345–360.
34. Golbach, F.; Kootstra, G.; Damjanovic, S.; Otten, G.; van de Zedde, R. Validation of plant part measurements using a 3D reconstruction method suitable for high-throughput seedling phenotyping. *Machine Vision and Applications* **2016**, *27*, 663–680.