IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received September 17, 2019, accepted October 4, 2019. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2019.2946369

Fruit Localization and Environment Perception for Strawberry Harvesting Robots

YUANYUE GE[1], YA XIONG[1], GABRIEL LINS TENORIO[2], AND PÅL JOHAN FROM[1]

[1]Faculty of Science and Technology, Norwegian University of Life Sciences, 1422 Ås, Norway
[2]Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22451-900, Brazil

Corresponding author: Yuanyue Ge (yuge@nmbu.no)

This work was supported by the Research Council of Norway, FORNY2020, under Project 2962020.

ABSTRACT This work presents a machine vision system for the localization of strawberries and environment perception in a strawberry-harvesting robot for use in table-top strawberry production. A deep convolutional neural network for segmentation is utilized to detect the strawberries. Segmented strawberries are localized through coordinate transformation, density base point clustering and the proposed location approximation method. To avoid collisions between the gripper and fixed obstacles, the safe manipulation region is limited to the space in front of the table and underneath the strap. Therefore, a safe region classification algorithm, based on Hough Transform algorithm, is proposed to segment the strap masks into a belt region in order to identify the pickable strawberries located underneath the strap. Similarly, a safe region classification algorithm is proposed for the table, to calculate its points in 3D and fit the points onto a 3D plane based on the 3D point cloud, so that pickable strawberries in front of the table can be identified. Experimental tests showed that the algorithm could accurately classify ripe and unripe strawberries and could identify whether the strawberries are within the safe region for harvesting. Furthermore, harvester robot's optimized localization method could accurately locate the strawberry targets with a picking accuracy rate of 74.1% in modified situations.

INDEX TERMS Robotics and automation, strawberry harvester, machine vision, environment perception.

I. INTRODUCTION

Machine vision is an essential element in agricultural robots. Before the development of deep learning techniques, traditional image processing methods were used, such as methods based on color thresholding, however these were not able to adapt to changing agricultural environments [1]–[3].

Deep Convolutional Neural Networks (CNN) have greatly improved the performance of image processing, particularly since the emergence of AlexNet, proposed by Krizhevsky *et al.* [4] and the numerous other detection CNN subsequently developed, some of which have been utilized for the detection of crops and fruits. Examples of such networks include You Only Look Once (YOLO), proposed by Redmon *et al.* [5], Single Shot Detector (SSD), proposed by Liu *et al.* [6] and the Region-based Convolutional Neural Network (Faster R-CNN), proposed by Girshick [7]. Sa *et al.* [8] utilized Faster R-CNN in the detection of sweet peppers, mangoes, strawberries and other fruit while Bargoti and Underwood [9] adopted the same network to detect apples and mangoes, further improving its detection performance through data augmentation.

Besides object detection, segmentation CNNs have also been adopted for other applications in agriculture. Popular semantic segmentation networks include Fully Convolutional Network (FCN) [10], SegNet [11], DeepLab [12] and U-net [10]. Popular instance segmentation networks include Sharp Mask [13] and Mask R-CNN [14]. Bargoti and Underwood [15] utilized a semantic segmentation network to detect apples and estimate the yield. In addition, Yu *et al.* [16] utilized Mask R-CNN [14] for strawberry detection and similarly, Gonzalez *et al.* [17] used the same network for blueberry detection. While detection and segmentation networks have been widely used for the detection and counting of fruit, their applications in fruit harvesting have been rarely reported. Most of these methods focused on image analysis, thus were not applied to a specific agricultural machine system.

In order to achieve the efficient and reliable picking of the objects, they need to be localized after detection.

The associate editor coordinating the review of this manuscript and approving it for publication was Kun Mean Hou.

Different methods based on different cameras have been used for the localization of fruits and other agricultural crops. These include the use of stereo cameras, depth cameras or single camera with extra assumptions.

Mehta and Burks [18] localized citrus fruits using a fixed monocular camera. Xiong *et al.* [1] used a single RGB (Red, Green, Blue) camera for weed localization, based on the assumption that the distance between the camera and the weed plane was fixed.

Single camera techniques are simple but limited in their depth determination and, therefore, much work has been done on the development of multiple camera systems. Font *et al.* [19] presented a stereo camera system for apple and pear localization. Mehta and Burks [20] investigated the fruit localization problems using multiple cameras based on the assumption that the target had been matched successfully. Similarly, Ji *et al.* [21] used stereo matching for the localization of apple branches.

Many agricultural robots use an RGB-D (RGB-Depth) camera for detection and localization because of its simplicity. Wang *et al.* [22] used an RGB-D camera for the detection and fruit size estimation of mangoes. Vitzrabin and Edan [23] proposed a detection method for sweet peppers using an RGB-D camera, and Xiong *et al.* [3] developed a strawberry harvester using an RGB-D camera for the detection and localization of the fruits. In this paper, we used an RGB-D camera for object detection and localization.

Environment perception or ambient awareness is crucial for agricultural robots, to ensure safe interaction between the robot and humans, the surrounding environment and other objects. Reina *et al.* [24] integrated Light Detection And Ranging (LiDAR) and imaging for the environment awareness of outdoor vehicles. Similarly, the same researchers [25] developed a multi-sensor system that integrates stereo-vision, LiDAR, radar and thermography, for the ambient awareness of agricultural vehicles in crop fields. They also [26] used RGB-D images to sense obstacles in outdoor environments in the navigation of rough terrain mobile robots. Indeed, the environment perception system is most commonly used for vehicle navigation, the conditions of which are markedly different to those for a strawberry picking robot on a strawberry farm. In order to ensure safe picking operations, it is necessary for the robot to detect the environment directly surrounding the target strawberries.

In the development of various strawberry harvesters, some have adopted machine vision systems based on color thresholding methods [2], [3], [27], utilizing the color differences to distinguish between ripe strawberries and other strawberries and plants. Some machine vision systems have been designed to detect the strawberry peduncle as they work with a scissor-like cutter to cut the peduncle [28]–[30]. These systems apply color thresholding to first detect the strawberry and then detect the peduncle of the strawberry by identifying a certain region above the strawberry. However, as mentioned above, this color-based image processing is not able to adapt to changing environments [3].

Traditional feature learning methods have most typically been used for learning the different shapes of strawberries [31] and deep learning techniques for object detection and segmentation have shown results in the detection of strawberries [8], [16], [32]. However, these work have focused on image processing and, as previously mentioned, when integrated with a real strawberry harvester, the accurate localization of the strawberries and maintenance of the safe picking operations are essential and are, therefore, the main focus of this paper.

Specially, we aim to solve the localization and collision problems frequently encountered during table-top picking for the strawberry harvester. The following highlights are presented in this paper:

- We utilize the deep learning network for instance segmentation to detect the target strawberries. Based on the detection results, we propose a localization method based on points clustering and location approximation algorithms.
- We raise the potential collision problems for manipulators in table-top strawberry farming. We solve this problem by proposing environment perception algorithms that can identity a safe manipulation region and the strawberries within this region. We propose the safe region classification method for the strap in a 2D image and the table in 3D point cloud to identify the pickable strawberries that are located underneath the straps as well as the pickable strawberries in front of the table.
- The methods for localization and environment perception were implemented and evaluated on our strawberry harvesting robot in the farm conditions, thus providing a reference for machine vision systems for localization and environment perception for similar harvesting robots.

II. OVERALL SYSTEM DESIGN

Our strawberry picking robot conducts static picking, in which it stops and processes the input image before issuing a command to the robot control system. Therefore, when the robot is static, the RGB and depth image acquired from the camera module is utilized for the computation of localization and environment perception in the machine vision system.

The overall architecture of the proposed machine vision system is shown in Fig. 1. Instance segmentation network Mask R-CNN was utilized to detect our targets, including strawberries, strap and table. Thereafter, the detected strawberries undergo safe operation checking in 2D imaging, coordinate transformation, a 3D location approximation algorithm and safe operation checking in 3D space, to obtain the final 3D strawberries' locations within the safe manipulation region, thus achieving safe and efficient picking.

The proposed environment perception algorithms include defining the safe manipulation region in 2D image according to the locations of the strawberries and strap, and defining the safe manipulation region in 3D according to the locations of the strawberries and table.

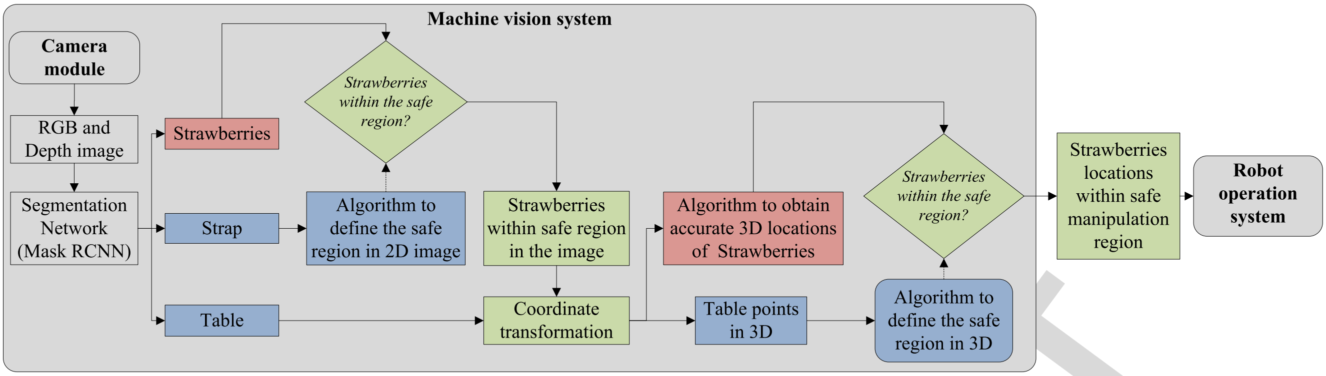

FIGURE 1. Overall architecture diagram.

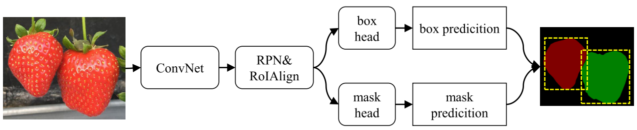

FIGURE 2. Mask R-CNN for strawberry fruits detection and segmentation.

In Fig. 1, the procedures related to strawberry localization are highlighted in red, while those related to environment perception are highlighted in blue. These two objectives coordinate with each other to finalize the positions of strawberries within the safe region, therefore the procedures relating to both objectives are highlighted in green. The detailed localization and perception algorithms will be described in the following sections.

III. INSTANCE SEGMENTATION AND LOCALIZATION

A. FRUITS DETECTION AND SEGMENTATION

Mask R-CNN [14] was used for the detection and segmentation of fruits, tables and straps. Mask R-CNN is a deep neural network that can generate both the bounding box and the masks for each instance, as can be seen in Fig. 2. ResNet101 was used as the base convolutional neural network for feature extraction.

As described above, there are several networks available for object detection that are fast, accurate and well suited for fruit counting and yield estimation [5]–[7]. However, our goal is to estimate the fruit location in 3D space as accurately as possible. In this case, segmentation can provide more detailed information and is thus more appropriate for localization, since the segmented masks only contain the pixels of the targets whereas bounding boxes additionally include pixels of other objects. To sum up, the instance segmentation method was used because it can generate pixel-level segmentation for each object.

Four target groups were classified, namely ripe strawberries, raw strawberries, straps and tables. The ripe strawberries are, of course, the harvester's target, while the tables and straps present potential collision problems with the gripper while in manipulation and are, therefore, also objects that should be detected. Detailed discussion about strap and table detection will be presented in the next section.

Three examples of the detection and segmentation results are provided in Fig. 3. Fig. 3 (a) shows the input images and Fig. 3 (b) displays the detection and segmentation results, including bounding boxes, masks and class names, while Fig. 3 (c) shows the colorized segmented pixel-level masks, with each color representing a different object.

B. COORDINATE TRANSFORMATION FOR SEGMENTED STRAWBERRIES

Through image processing, several masks were created for the strawberries, in which one mask represented a detected target. The masks were de-projected into 3D points, representing the 3D positions of the targets in the camera frame C. The workflow of the coordinate transformation is shown in Fig. 4. The masks were extracted from the detected results and the depth image was aligned to the RGB coordinate system. The depth value was then obtained by matching the aligned depth image with the corresponding mask results. The coordinates were transformed from the image frame I to the RGB camera optical frame C using the intrinsic parameters of the RGB-D camera.

Examples of the coordinate transformation process and its results can be seen in Fig. 5. The first and second columns are the colorized detected masks and the corresponding depth images, respectively. The third column is the visualization of transformed points marked by 3D bounding boxes in the point cloud. The detected masks contain the unripe strawberries but only the positions of the ripe strawberries were selected and sent to the harvester. Therefore, the third column shows the 3D bounding boxes of the ripe strawberries.

C. TARGET LOCATION APPROXIMATION METHODS

1) POINTS CLUSTERING

In this harvesting system, once the 3D positions of the targets are obtained, the machine vision system needs to send the positions of all strawberries to the manipulation system. However, it was found that the raw points transformed from the masks were not sufficiently accurate.

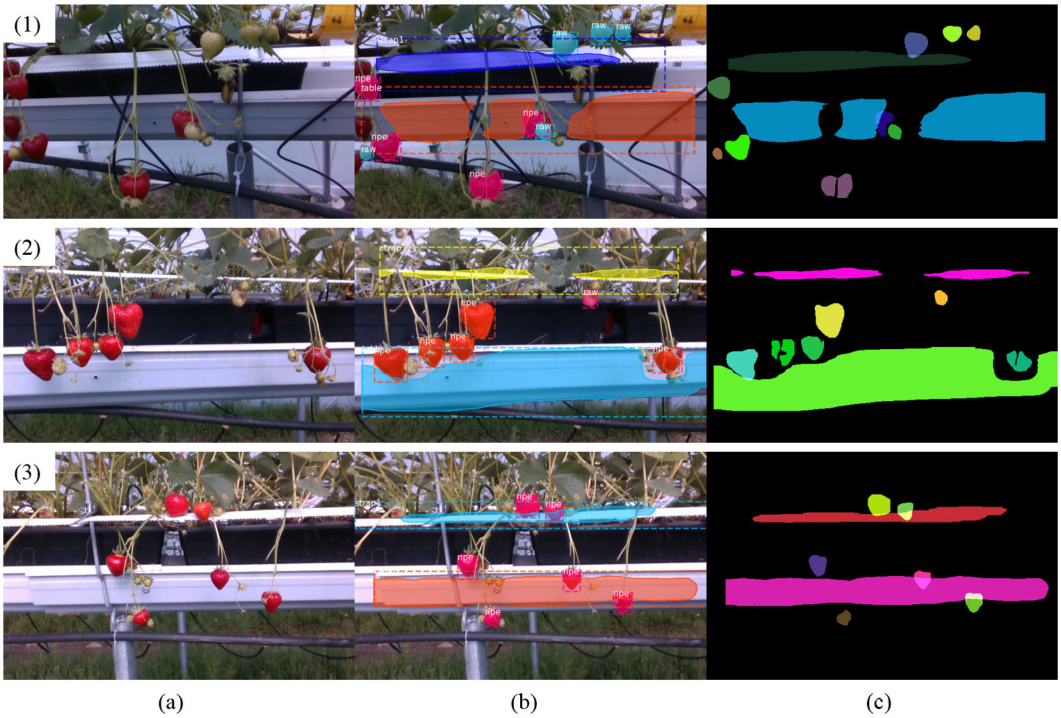

FIGURE 3. Detection and segmentation results. (1)-(3) are three examples. (a) shows the input images; (b) displays the visualized segmentation results on the input image; (c) shows the colorized segmented pixel-level masks.

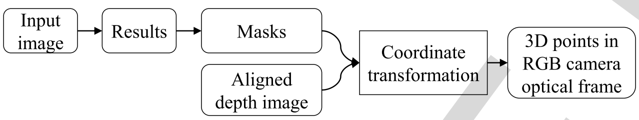

FIGURE 4. Workflow of the coordinate transformation.

Therefore, post-processing procedures were implemented on the raw points to obtain a point-set that could better represent the target's real position.

The inaccuracy of the transformed points was caused by several factors. For example, the target points could be projected to the background scene due to inaccurate sensing from the depth camera, such as the example shown in Fig. 6 (a). Another factor was noise from the adjacent objects and, in addition, there may have been inaccurate segmentation of the masks from the Mask R-CNN.

Therefore, a clustering algorithm was utilized to screen out irrelevant or noisy points. Density-Based Spatial Clustering (DBSC) of applications with a noise algorithm [33] is a method that in which group points can be closely packed together. By setting a threshold distance to measure core samples and a parameter of a minimum number of points that can be a cluster, the less dense points and noises could be removed. Fig. 6 shows three examples of points before and after clustering, enclosed in the bounding boxes. The noises marked in the figure, can be filtered through this clustering method. Fig. 6 (a) shows an example of a strawberry edge sticking to the background, while 6 (b) and (c) show the examples of noises caused by adjacent objects.

2) TARGET POSITION OPTIMIZATION

The 3D bounding boxes of target strawberries in the RGB camera optical frame were sent to the manipulator. The raw points obtained after clustering and the bounding box that encloses the region of the points is shown in Fig.7 (a), in which it is evident that the bounding box can only represent a portion of a strawberry. The surface of the target that faces towards the camera is sensed better than other surfaces as the RGB-D camera uses a projection method to obtain 3D points. In the table-top scenario, if the camera angle is that of the front view, the lengths in the x and z dimensions of a strawberry are almost the same. Therefore, in order to localize the targets more accurately, we used the dimensions detected in the x axis (representing the surface towards the camera) to represent those in the z axis. Fig.7 (b) shows the strawberry points and the refined bounding box.

D. WORLD COORDINATE TRANSFORMATION

The camera module enabled the location of the 3D coordinates of the fruit in the camera optical frame C, so it was necessary to convert the locations from the camera frame C into the arm frame W. The relationship between the different frames is shown in Fig. 8, in which S represents the strawberry, C the camera frame, W the arm frame and B the chess board frame.

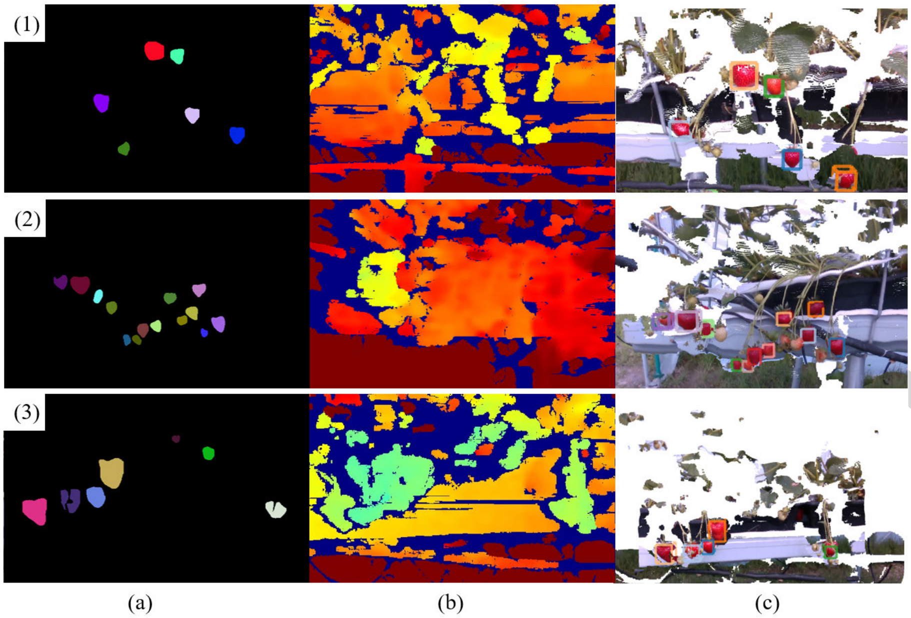

FIGURE 5. Examples of coordinate transformation for strawberries: (a) detected masks, with each color representing a detected strawberry; (b) is the colorized depth image; (c) localization results visualized in point cloud using bounding boxes.

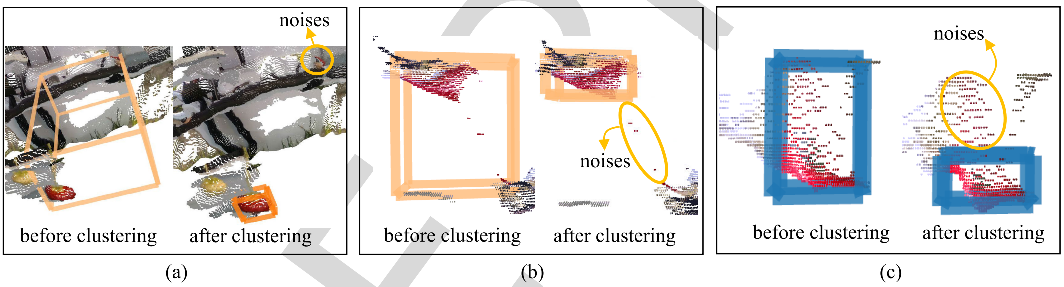

FIGURE 6. Three examples of clustering of strawberry points.

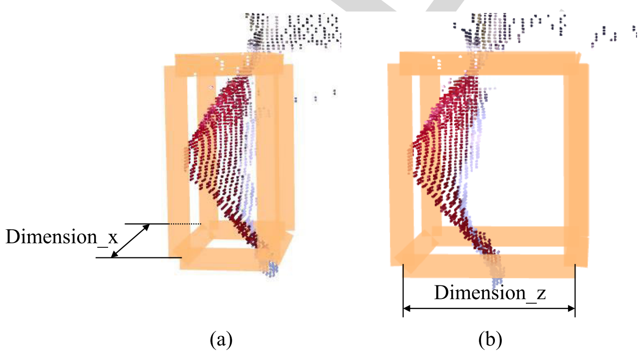

FIGURE 7. Position optimization: (a) the bounding box of a strawberry that encloses the filtered points; (b) the optimized bounding box and corresponding strawberry points.

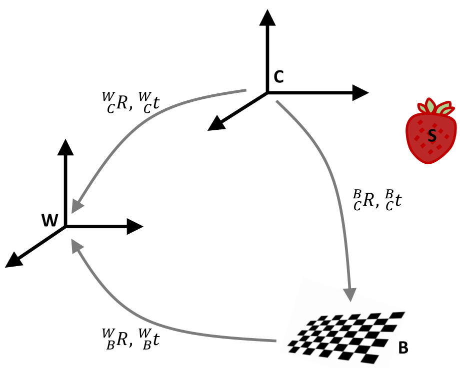

FIGURE 8. Frames for world coordinate transformation.

Let ${}^W S$ be the location of the strawberry S with respect to the arm frame W, and ${}^C S$ be defined as the location of strawberry S location in the camera frame. The coordinate transformation of strawberries from camera frame to arm frame can be expressed as follows:

$$ {}^W S = {}^W_C R * {}^C S + {}^W_C t \tag{1} $$

where ${}^W_C R$ and ${}^W_C t$ are the rotation matrix and translation vector from the camera frame C to the arm frame W.

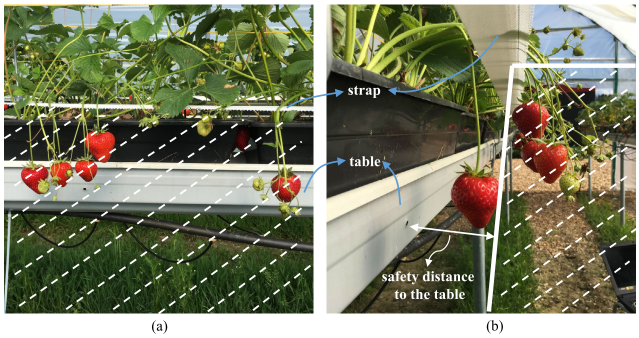

FIGURE 9. The safety manipulation region for the strawberry picking robot. (a) is a front view with the safety region marked by white dash line; (b) is a side view with the safety region marked by white dash line.

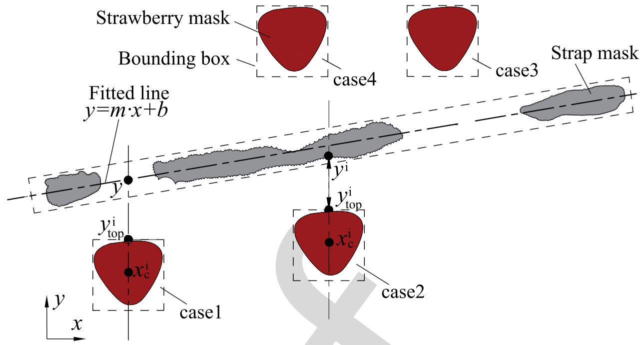

FIGURE 10. Schematic of safety solution calculation for the straps: (1) using method 1, case 1, case 2 and case 4 would be considered successful, while case 3 would be a failure; (2) using method 2, all cases would be considered successful.

The ${}^{B}_{C}R$, ${}^{B}_{C}t$ shown in Fig. 8 can be obtained through camera calibration while ${}^{W}_{B}R$, ${}^{W}_{B}t$ are known parameters. Based on these two sets of parameters, ${}^{W}_{C}R$ and ${}^{W}_{C}t$ can be obtained.

IV. ENVIRONMENT PERCEPTION

A. PROBLEM DEFINITION

It is necessary for the strawberry harvester to sense its environment in order to make predictions and plan for the manipulation. Therefore, the scene must be segmented and objects that could cause potential damage must be localized.

During the experiments, the manipulator collided with the table or strap when the strawberries were either too close to the table or above the strap. Therefore, we used the segmentation network to detect the strap and table and make estimations about whether or not a target strawberry was located within the safe manipulation region. The regions marked by white dash lines in Fig. 9 represent the safe safety region for the manipulation. Fig. 9 (a) is a front view of the scene, in which the safe region is below the strap, while Fig. 9 (b) shows a side view showing the safe region below the strap and a safety distance from the table. Strawberries should, therefore, be picked in the safe region.

B. SAFETY SOLUTIONS FOR THE STRAPS

An important output obtained by the Mask R-CNN model was the strap masks. The strap above the strawberry table is used to support the strawberries plant during growth, making fruit easier to harvest and also preventing the stems from breaking. Most ripe strawberries hang underneath the straps, however some can be found above the straps, which may be dangerous for the gripper during harvesting. In this section, we introduce two methods by which strawberry positions can be identified in relation to the strap.

1) METHOD 1: ORIGINAL MASKS

In order to classify the strawberries that are on or above the straps, the top positions (y^i_{top}) and the horizontal centroids (x^i_c) of the strawberries bounding boxes are first calculated, as shown in Fig. 10. Thereafter, for each strap mask region of non-zero pixels, x^i_c is applied to obtain all the vertical coordinates $\mathbf{y}^i$ from the masks. Next, y^i_{top} is compared to the minimum value of $\mathbf{y}^i$, which is used to represent the strap position, and assigned as dangerous if the strawberries are above the strap and safe if the strawberries are below the strap.

We observed, however, that this method was not always sufficiently precise, as there were some situations in which corrupted segmented straps were obtained, such as case 3 shown in Fig. 10. In this case, the calculation method was not applicable to the strawberries that did not have strap masks below and, therefore, case 3 may be considered a failure using this method.

2) METHOD 2: RECTIFIED MASKS

To solve the above mentioned problems arising in method 1, first, the Canny Edge Detection algorithm proposed by Canny [34] was applied to ascertain all of the edge points of a segmented strap. Thereafter, we sequentially applied the Probabilistic Hough Transform algorithm proposed by Kiryati *et al.* [35], which uses a random subset from the edge detector to obtain multiple lines in the image, including their starting and ending coordinates. All these coordinates were then used to calculate the line equation ($y = m \cdot x + b$) that best interpolates all the points by using least squares. The bounding box that enclosed all the strap masks, marked by the dash line in Fig. 10, was determined by the width of the strap and the fitted line. As shown in Fig. 10, to verify whether strawberries are above or below the straps and assign a warning sign (dangerous or safe) to each fruit, x^i_c is applied to the line equation to obtain the y and compare it to the $y^i_{\text{top}} + \textit{threshold}$. This *threshold* is a value obtained through the original segmented mask to determine the safe manipulation region between the line and the position of the top of the fruit. As shown in Fig. 10, all cases were defined correctly using this method.

Comparative visual results for the two methods described above, the safety solution containing the original strap segmentation and the rectified strap segmentation, are shown in Fig. 11. The images Fig. 11 (a) presents the original images, while the images in Fig. 11 (b) show the results of the first method and the images in Fig. 11 (c) show

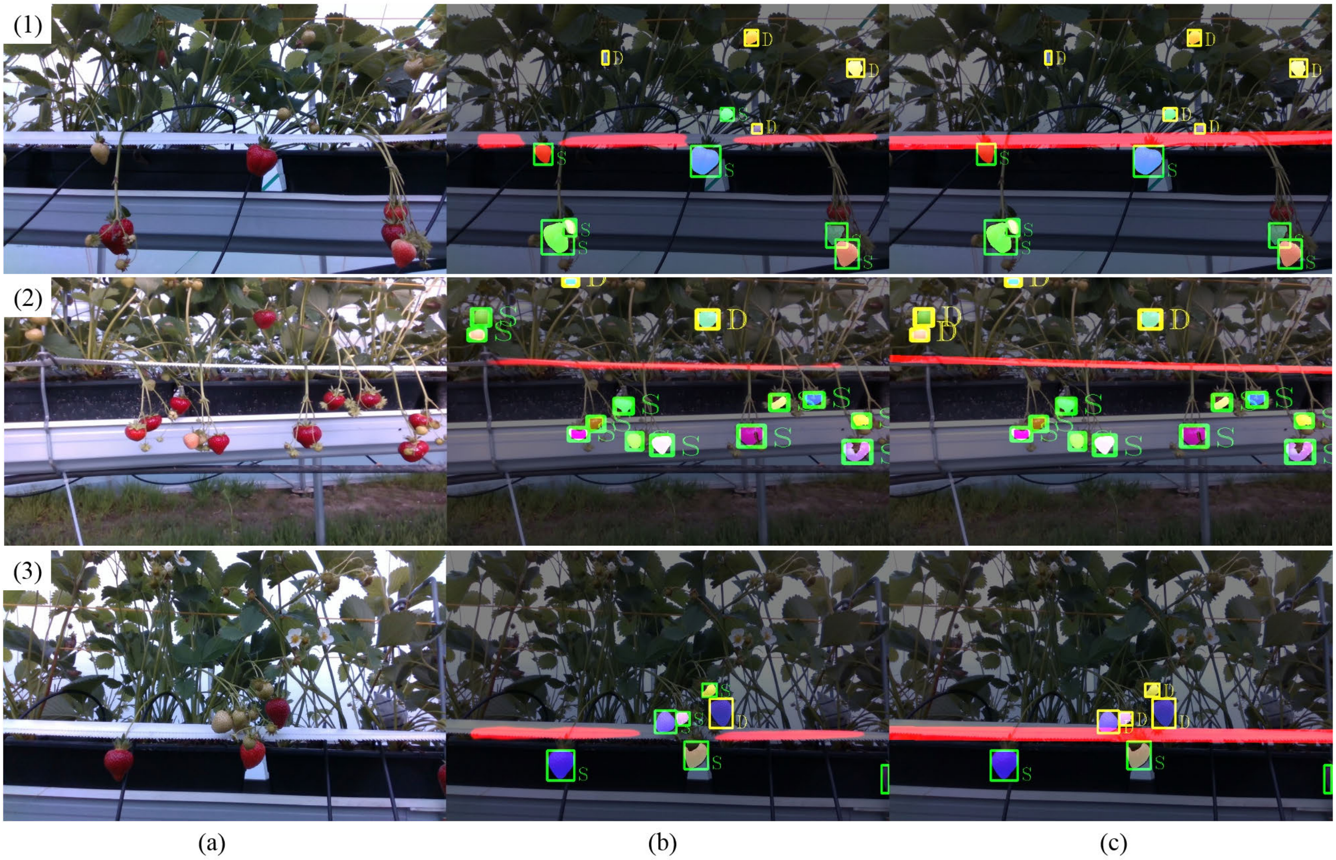

FIGURE 11. **Visual results of the safety solution for the original strap segmentation and the rectified strap segmentation: (a) original images (1,2,3); (b) the image results of the first method; (c) image results of the second method; The green and yellow bounding boxes indicate, the safe (S) and the dangerous (D) warning signs.**

the results of the second method. The green and yellow bounding boxes indicate, the safe (S) and the dangerous (D) warning signs, respectively. It is evident from these images that the visual results obtained through the first method could not correctly classify as dangerous the strawberries above the corrupted regions of the strap masks. However, with the second method, all the fruits were classified successfully.

C. SAFETY SOLUTION FOR THE TABLE

The picking robot needs to know the specific 3D location of the table in order to identify the proximity of a strawberry. The same clustering method was used for the table 3D points. The detected table masks and corresponding 3D points for table can be seen in Fig. 5.

In order to represent a table's complete position, we fitted a 3D plane to the detected 3D points of the table. A plane in 3D space can be determined by defining a point $p_0 = (x_0, y_0, z_0)$ on the plane and a normal vector $n = (a, b, c)$ that is perpendicular to the surface. The surface $p = (x_p, y_p, z_p)$ can be represented by $n \cdot (p - p_0) = 0$.

We used the centroid of the points as p_0. Then we created a moment of inertia tensor and used singular value decomposition to obtain the normal vector $\boldsymbol{n}$ of the plane.

The distance between the detected strawberry center p_s and the table surface plane p could then be calculated. A line $l = (x_l, y_l, z_l)$ passing through point p_s and perpendicular to the table plane can be represented by $l = k * n + p$. The intersection point p_i between the line and the plane satisfies both equations as follows:

$$\begin{cases} l = k * n + p_i \\ n \cdot (p_i - p_0) = 0 \end{cases} \tag{2}$$

Thus the value of k and the exact position of p_i were obtained. The distance between p_i and p_s was calculated and used to ascertain whether or not a strawberry is within the dangerous distance to the table of strawberry trays.

The results of the detection and segmentation results of table are presented in Fig.12 (a). The detected coordinates in the image can be obtained from the masks and transformed to the camera optical frame with the aligned depth image. The fitted plane is marked in green in Fig.12 (b) and Fig.12 (c). Fig.12 (c) also shows the point cloud and the

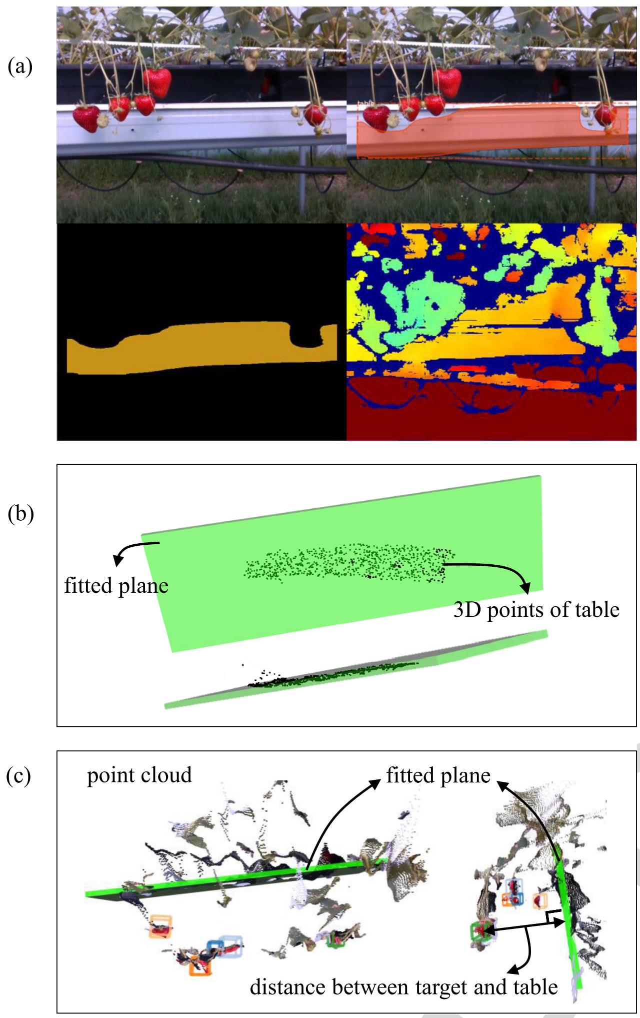

FIGURE 12. **Coordinate transformation and surface fitting for table: (a) the input image, visualized segmentation results in the input image, detected mask and corresponding depth image; (b) the transformed 3D points (highlighted in black) and the fitted 3D plane (highlighted in green); (c) point cloud with corresponding fitted table plane and detected strawberries.**

detected strawberries, as well as the distance between the target and the table.

D. STRAWBERRIES IN THE SAFE MANIPULATION REGION

The coordinates of detected strawberries were compared with the positions of the strap and table, to ascertain whether a strawberry was within the safe region. The algorithm for the position checking sequence can be seen in Algorithm 1.

The entire process can be concluded within the following three main steps. First, the positions of the strawberry and strap are compared within the 2D image, disregarding any strawberries above the strap. Second, the positions of the strawberry and the table are compared in the 3D space in the RGB camera's optical frame. The remaining strawberries and the table are also compared in 3D space, with those strawberries close to the table screened out by the pre-defined safety distance. In the third and final step, only the strawberries below the strap and outside the safety distance to the table are selected.

Algorithm 1 Ascertain Whether Strawberries Are Within the Safe Region

Result: coordinates of strawberries in safe manipulation region
pre-processing: 2D line fitting for the strap and 3D plane fitting for the table. ;
for *every detected strawberry* **do**
  comparing the strawberry position with strap line and table surface;
  if *the strawberry is above the strap* **then**
    remove the position of this strawberry target;
  else if $Dist2T < Dist\_safe\_limit$ **then**
    remove the position of this strawberry target;
  else
    keep the position of this strawberry target;
  end
end

TABLE 1. **Evaluation results of detection method.**

Class	Confidence	Precision	Recall	F1	AP
ripe strawberry	0.7	0.91	0.95	0.93	0.90
	0.8	0.95	0.93	0.94	
	0.9	0.97	0.92	0.94	
unripe strawberry	0.7	0.85	0.83	0.84	0.72
	0.8	0.89	0.84	0.86	
	0.9	0.93	0.86	0.89	

V. EXPERIMENTS

A. EVALUATIONS OF DETECTION METHOD

The metrics used to evaluate the detection results include precision, recall, F1 score and Average Precision(AP), as defined in Eq. 3, below. A total of 120 images were used to evaluate the detection method and the number of True Positive (TP) and False Positive (FP) were recorded. Three confidence values, ranging from 0.7-0.9, were set to compute the precision, recall, F1 score and AP. The results are shown in Table 1, in which it can be seen that ripe strawberries had a higher rate of detection accuracy. It was evident that from the annotation process that the ripe strawberries are easy to define while unripe strawberries are more difficult as they undergo a long growth stage from young, small strawberries to partially ripe strawberries. This could be confusing to the detection network.

$$\begin{cases} precision = \dfrac{TPs}{TPs + FPs} \\ recall = \dfrac{TPs}{GTs} \\ F1 = \dfrac{2 \times precision \times recall}{precision + recall} \\ AP = \int_0^1 p(r)\, dr \end{cases} \tag{3}$$

TABLE 2. Confusion metrics for the safety solution methods of straps: Method 1 (original masks) and Method 2 (rectified masks).

		Predicted	
		Dangerous	Safe
Actual	Dangerous (Original)	80	60
	Safe (Original)	8	270
	Dangerous (Rectified)	117	4
	Safe (Rectified)	9	288
Overall accuracy		Original: 83.7% Rectified: 96.9%	

TABLE 3. Confusion matrix for the safety solution of table.

		Predicted	
		outside Dist_danger	within Dist_danger
Actual	outside Dist_danger	98	2
	within Dist_danger	1	11
Overall accuracy: 97.3%			

B. EXPERIMENTS OF SAFETY SOLUTION FOR THE STRAPS

The performance of the two safety solution methods for the straps were evaluated, using test images containing a total of 418 strawberries. It is relevant to mention the strawberries were most commonly situated below the strap, so the warning sign classification was highly unbalanced. Confusion metrics for both methods are presented in Table 2, in which it is evident that the results for the method involving the original masks show high classification errors for the dangerous warning sign class. Some of the Dangerous classes were classified as Safe mainly due to the corrupted regions of the strap masks. However, after rectifying the masks, this error was mitigated and the overall accuracy results were improved from 83.7% to 96.9%.

In both methods, the inaccurate classifications (Safe classified as Dangerous) were due to poor segmentation as well as inaccurate line equations.

C. EXPERIMENTS OF SAFETY SOLUTIONS FOR THE TABLE

The safety solutions for the table were evaluated using the RGB images, aligned depth images and point cloud. The RGB and depth images were used for obtaining detection and localization results while the ground truth was obtained by manually measuring the distance between the target and the table in the point cloud. The safety distance was set to 10 cm based on reasonable practical experience. Twenty sets of the collected data with 112 strawberries were tested and the classification results are shown in the confusion matrix in Table 3. Similar to straps results, significantly fewer strawberries were found in the dangerous region than in the safe region. The overall accuracy was 97.3%.

The accuracy of the plane fitting was based on accurate detection and localization of the table. Therefore, the evaluations were primarily based on the assumption that the table had been correctly detected. Should the points not sufficiently accurate, the resulting fitted plane may not be well aligned

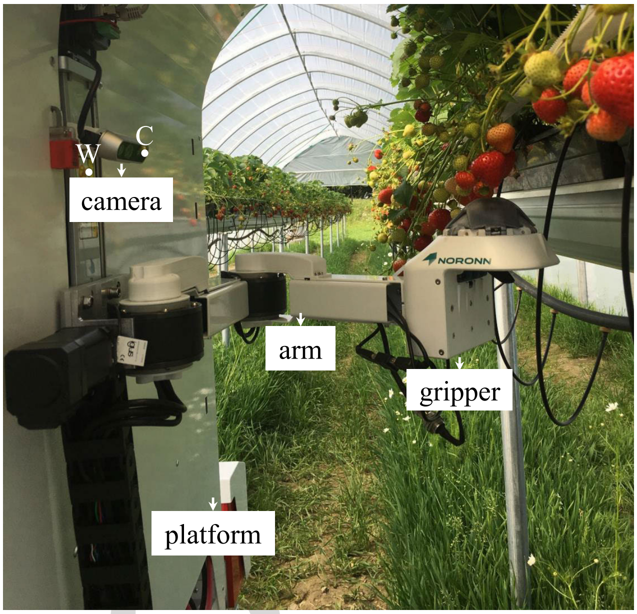

FIGURE 13. Strawberry harvester, developed by Noronn AS, including the platform, camera, robotic arm and gripper: W and C represent the origins of arm and camera frame, respectively.

TABLE 4. Timing of the machine vision system.

	detection (s)	transformation (s)	others (s)	total (s)
average	0.62	0.20	$4.0e-6$	0.82
st_dev	0.02	0.04	$1.5e-6$	0.04

to the real table. Because the aim of the algorithm is to accurately identify the strawberries within the safe manipulation region, the confusion matrix was used that would reflect related failures.

D. EVALUATION OF LOCALIZATION ON THE HARVESTING ROBOT

We tested the strawberry detection and localization method on our strawberry harvester (developed by Noronn AS). This harvester comprises a vehicle platform, a camera, a robotic arm and a gripper for picking strawberries [3], [36], as shown in Fig.13. A GPU (GTX 1060, NVIDIA, USA) was used for running the machine vision and manipulation control systems. The average processing time for one image frame, including running the detection network, coordinate transformation and other computations was 0.82s, as can be seen in Table 4. The time is an average of 119 image frames with a resolution of 640 × 480. The average times and their standard deviations for processing the detection, coordinate transformation (including strawberries and table points) and other computations are listed separately in Table 4.

The successful picking rates of the localization method based on raw points (method 1) and the bounding box optimization (method 2) were compared using the same scenarios, in which the cutting action was disabled so that the gripper swallowed the strawberry, moved down and went

TABLE 5. Picking success rate with the localization method.

test No.	Number of detected	Number of swallowed	
		method1	method2
1	4	3	4
2	1	0	1
3	5	4	4
4	4	2	4
5	1	1	1
6	4	4	4
7	8	3	5
8	7	2	4
9	5	2	3
10	6	3	3
11	8	4	6
12	5	2	4
Accuracy		51.8 %	74.1%

to the next strawberry. Each successful swallowing was considered as a successful picking.

The tests were conducted in modified situations, including those in which the strawberries were isolated and those in which ripe and raw strawberries were hanging adjacent to each other. In this test, the Rumba variety of strawberry was used, and the number of successfully detected and successfully swallowed strawberries of 12 trials are recorded in Table 5. The test of different growing situations can also be found in [36], in which the various harvesting failure cases were introduced. The picking rate in this paper is lower than that in [36], because in this test the variety of strawberry is more challenging for picking and the tests were conducted with one attempt of picking.

The picking rates for the two localization methods were obtained by dividing the swallowed strawberries by the number of detected strawberries. Method 1 in Table 5 indicates localization based on raw points, while method 2 indicates the optimized localization method. It can be seen that the optimized localization method achieved a success rate of 74.1% in the modified environment, while the localization based on raw points achieve a successful picking rate of 51.8%.

VI. CONCLUSION

This work proposed a localization method and environment perception algorithms for strawberry harvesting robots. The localization method was based on the segmented masks of a deep convolutional neural network and depth images from an RGB-D camera. To increase localization accuracy, density based point clustering was used to segment and remove noise points in the 3D point cloud. The table and strap were detected and located using the same network, and their locations were compared with the positions of strawberries in order to identify whether the strawberries were within the safe manipulation region. The position comparison between the target strawberries and the strap was based on the line fitting using the Hough Transform algorithm, while the position comparison between strawberries and the table was based on a 3D plane fitting. The test results showed that the optimized localization method can accurately localize targets, with an accurate picking rate of 74.1% in modified situations. The overall accuracy rates for the strap and table safety identifications were 96.9% and 97.3%, respectively.

This work investigated the challenges of localization based on deep learning segmentation networks. It also raised the problem of environment perception in harvesting and provided methods for detecting the danger objects for the harvester and classifying the safe manipulation region.

In future work, the localization algorithm could be further optimized and adopted to suit more complex situations, such as occluded and unusual hanging positions of the strawberries.

REFERENCES

[1] Y. Xiong, Y. Ge, Y. Liang, and S. Blackmore, "Development of a prototype robot and fast path-planning algorithm for static laser weeding," *Comput. Electron. Agricult.*, vol. 142, pp. 494–503, Nov. 2017.

[2] S. Hayashi, S. Yamamoto, S. Saito, Y. Ochiai, J. Kamata, M. Kurita, and K. Yamamoto, "Field operation of a movable strawberry-harvesting robot using a travel platform," *Jpn. Agricult. Res. Quart., JARQ*, vol. 48, no. 3, pp. 307–316, Jul. 2014.

[3] Y. Xiong, C. Peng, L. Grimstad, P. J. From, and V. Isler, "Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper," *Comput. Electron. Agricult.*, vol. 157, pp. 392–402, Feb. 2019.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.

[7] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

[8] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. Mccool, "DeepFruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, Sep. 2016.

[9] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2017, pp. 3626–3633.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2017.

[13] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár, "Learning to refine object segments," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 75–91.

[14] K. He, G. Gkioxari, and P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[15] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in Apple orchards," *J. Field Robot.*, vol. 34, no. 6, pp. 1039–1060, Sep. 2017.

[16] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104846.

[17] S. Gonzalez, C. Arellano, and J. E. Tapia, "Deepblueberry: Quantification of blueberries in the wild using instance segmentation," *IEEE Access*, vol. 7, pp. 105776–105788, 2019.

[18] S. S. Mehta and T. F. Burks, "Vision-based control of robotic manipulator for citrus harvesting," *Comput. Electron. Agricult.*, vol. 102, pp. 146–158, Mar. 2014.

[19] D. Font, T. Pallejà, M. Tresanchez, D. Runcan, J. Moreno, and D. Martínez, M. Teixidó, and J. Palacín, "A proposal for automatic fruit harvesting by combining a low cost stereovision camera and a robotic arm," *Sensors*, vol. 14, no. 7, pp. 11557–11579, Jun. 2014.

[20] S. S. Mehta and T. F. Burks, "Multi-camera fruit localization in robotic harvesting," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 90–95, 2016.

[21] W. Ji, X. Meng, Z. Qian, B. Xu, and D. Zhao, "Branch localization method based on the skeleton feature extraction and stereo matching for apple harvesting robot," *Int. J. Adv. Robotic Syst.*, vol. 14, no. 3, May 2017, Art. no. 1729881417705276.

[22] Z. Wang, K. B. Walsh, and B. Verma, "On-tree mango fruit size estimation using RGB-D images," *Sensors*, vol. 17, no. 12, p. 2738, Nov. 2017.

[23] E. Vitzrabin and Y. Edan, "Changing task objectives for improved sweet pepper detection for robotic harvesting," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 578–584, Jan. 2016.

[24] G. Reina, A. Milella, W. Halft, and R. Worst, "LIDAR and stereo imagery integration for safe navigation in outdoor settings," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot. (SSRR)*, Oct. 2013, pp. 1–6.

[25] G. Reina, A. Milella, R. Rouveure, M. Nielsen, R. Worst, and M. R. Blas, "Ambient awareness for agricultural robotic vehicles," *Biosyst. Eng.*, vol. 146, pp. 114–132, Jun. 2016.

[26] G. Reina, M. Bellone, L. Spedicato, and N. I. Giannoccaro, "3D traversability awareness for rough terrain mobile robots," *Sensor Rev.*, vol. 34, no. 2, pp. 220–232, Mar. 2014.

[27] S. Yamamoto, S. Hayashi, H. Yoshida, and K. Kobayashi, "Development of a stationary robotic strawberry harvester with a picking mechanism that approaches the target fruit from below," *Jpn. Agricult. Res. Quart., JARQ*, vol. 48, no. 3, pp. 261–269, Jul. 2014.

[28] S. Hayashi, K. Shigematsu, S. Yamamoto, K. Kobayashi, Y. Kohno, J. Kamata, and M. Kurita, "Evaluation of a strawberry-harvesting robot in a field test," *Biosyst. Eng.*, vol. 105, no. 2, pp. 160–171, Feb. 2010.

[29] Z. Huang, S. Wane, and S. Parsons, "Towards automated strawberry harvesting: Identifying the picking point," in *Proc. Annu. Conf. Towards Auto. Robotic Syst.* Springer, 2017, pp. 222–236.

[30] Y. Cui, Y. Gejima, T. Kobayashi, K. Hiyoshi, and M. Nagata, "Study on Cartesian-type strawberry-harvesting robot," *Sensor Lett.*, vol. 11, nos. 6–7, pp. 1223–1228, Nov. 2013.

[31] T. Ishikawa, A. Hayashi, S. Nagamatsu, Y. Kyutoku, I. Dan, T. Wada, K. Oku, Y. Saeki, T. Uto, and T. Tanabata, "Classification of strawberry fruit shape by machine learning," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2, pp. 463–470, May 2018.

[32] H. Habaragamuwa, Y. Ogawa, T. Suzuki, T. Shiigi, M. Ono, and N. Kondo, "Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network," *Eng. Agricult., Environ. Food*, vol. 11, no. 3, pp. 127–138, Jul. 2018.

[33] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96. Aug. 1996, pp. 226–231.

[34] J. Canny, "A computational approach to edge detection," in *Readings in Computer Vision: Issues, Problem, Principles, and Paradigms*. Amsterdam, The Netherlands: Elsevier, 1987, pp. 184–203.

[35] N. Kiryati, Y. Eldar, and A. M. Bruckstein, "A probabilistic Hough transform," *Pattern Recognit.*, vol. 24, no. 4, pp. 303–316, 1991.

[36] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," *J. Field Robot.*, vol. 36, pp. 1–23, Aug. 2019.

YUANYUE GE received the B.Sc. and M.Sc. degrees in vehicle engineering from China Agricultural University, Beijing, in 2013 and 2016, respectively, and the M.Sc. degree in applied mechatronic engineering from Harper Adams University, U.K., in 2016. She is currently pursuing the Ph.D. degree in agricultural robotics and machine vision with the Norwegian University of Life Sciences. Her research interests include agriculture robotics and machine vision.

YA XIONG received the B.Sc. and M.Sc. degrees in vehicle/mechanical engineering from China Agricultural University, Beijing, in 2013 and 2016, respectively, and the M.Sc. degree in mechatronic engineering from Harper Adams University, U.K., in 2016. He is currently pursuing the Ph.D. degree with the Agricultural Robotics and Laboratory Automation, Norwegian University of Life Sciences. He was a Visiting Ph.D. Student with the University of Minnesota, from May 2017 to August 2017. His research interests include agricultural robotics and laboratory automation, especially on manipulator design and its control.

GABRIEL LINS TENORIO received the B.Sc. degree in control and automation engineering and the M.Sc. degree in image processing, automation, and robotics from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Brazil, where he is currently pursuing the Ph.D. degree.

He was an AI Researcher with the Applied Computational Intelligence Laboratory (ICA) in partnership with Intel and Petrobras Research Center (Cenpes), from 2018 to 2019. He has two international publications in the area of deep learning, presented as a Conference Speaker. He participated for three consecutive years (July—2017–2019) in the research and development project at the Norwegian University of Life Sciences in the area of agricultural robotics. This project was supported by the UTFORSK Partnership Programme.

PÅL JOHAN FROM received the Ph.D. degree in modeling and control of complex robotic systems from the Norwegian University of Science and Technology.

Since 2010, he has been the Head of the Robotics Group, Norwegian University of Life Sciences, which has designed and built the Thorvald agricultural robot. He is currently a Professor of agri-robotics with the Norwegian University of Life Sciences and also with the University of Lincoln, U.K. He is also the CEO of saga robotics, which develops and commercializes the agricultural platform Thorvald. He has over 50 international publications in robotics and has written one book. He has also held a large number of peer-reviewed grants from various sources. These include both research grants and grants for commercialization.