Literature study on robotic tree harvesting and pruning

The Next Fruit 4.0 - Multifunctional robot deliverable 1

Authors | Pieter Blok, Jochen Hemming, Hitoshi Nishikawa and Menno Sytsma

Report WPR-OT-987

WAGENINGEN
UNIVERSITY & RESEARCH

Literature study on robotic tree harvesting and pruning

The Next Fruit 4.0 - Multifunctional robot deliverable 1

Pieter Blok
Jochen Hemming
Hitoshi Nishikawa
Menno Sytsma

Wageningen University & Research

This study was carried out by the Wageningen Research Foundation (WR) Business Unit Field Crops and was commissioned and financed, in part, by the TKI Topsector Tuinbouw & Uitgangsmaterialen, LWV20.131, The Next Fruit 4.0 (project number 3750435100).

WR is part of Wageningen University & Research, the collaboration of Wageningen University and Wageningen Research Foundation.

Wageningen, May 2022

Report WPR-OT-987

WAGENINGEN
UNIVERSITY & RESEARCH

Pieter Blok, Jochen Hemming, Hitoshi Nishikawa, Menno Sytsma, 2022. Literature study on robotic tree harvesting and pruning; The Next Fruit 4.0 - Multifunctional robot deliverable 1. Wageningen Research, Report WPR-OT-987.

This report can be downloaded for free at https://doi.org/10.18174/590172.

Abstract: This report contains three literature studies. In the literature study on pear detection methods the most relevant and recent publications on deep-learning based image analysis methods for fruit detection have been studied. Four software methods are popular: instance segmentation + 3D vertex normal estimation; extremely accurate instance segmentation; key-point detection and panoptic segmentation. The literature study robotic tree fruit harvesting groups the literature in methods for approaching the fruit/target, gripper requirements and manipulation patterns. For the literature study on machine vision systems for robotic pruning, a closer look is taken to literature describing the sensing system (mainly 2D or 3D cameras), data pre-processing (to create a high quality point clouds and to improve the accuracy) and key-point detection (to define pruning locations).

Keywords: robotic harvesting, robotic pruning, deep-learning, pears, red currant

© 2022 Wageningen, Stichting Wageningen Research, Wageningen Plant Research, Business Unit Field Crops, P.O. Box 16, 6700 AA Wageningen, The Netherlands; T +31 (0)317 48 07 00; www.wur.eu/plant-research

Chamber of Commerce no. 09098104 at Arnhem
VAT NL no. 8065.11.618.B01

Stichting Wageningen Research. All rights reserved. No part of this publication may be reproduced, stored in an automated database, or transmitted, in any form or by any means, whether electronically, mechanically, through photocopying, recording or otherwise, without the prior written consent of the Stichting Wageningen Research.

Stichting Wageningen Research is not liable for any adverse consequences resulting from the use of data from this publication.

Report WPR-OT-987

Contents

Summary

The main objective of the multifunctional robot case in the The Next Fruit 4.0 project is to expand the functionality of existing orchard robots as well as of orchard robots currently under development in parallel research projects. The focus of the work is on two topics, namely the development of a sensing system and a gripper for picking pears and on a sensing system, robot control and end-effector(s) for robotic pruning of fruit trees and red currant bushes.

In the literature study on pear detection methods, the most relevant and recent publications on deep-learning based image analysis methods for fruit detection, have been studied. Four software methods are popular: instance segmentation + 3D vertex normal estimation; extremely accurate instance segmentation; key-point detection and panoptic segmentation. It can be concluded that any of the four proposed methods can enable the detection of the pear and its peduncle. One method that stands out is the extremely accurate instance segmentation method.

The literature study robotic tree fruit harvesting groups the literature in methods for approaching of the fruit/target, gripper requirements and manipulation patterns. Reachability has been studied, but methods to improve this not. Little literature was found on robotic harvesting of pears. Research has been performed on what forces are needed to detach a fruit, but no examples have been found where it is also converted to controlling a robot to apply the same delicate movements as are performed by human harvesters.

In the literature study on machine vision systems for robotic pruning a closer look is taken to literature describing the sensing system (mainly 2D or 3D cameras), data pre-processing (to create a high quality point clouds and to improve the accuracy) and key-point detection (to define pruning locations).

1 Introduction

This report contains the literature studies of the multifunctional robot case study of the public-private-partnership project The Next Fruit 4.0. The main objective of the multifunctional robot case is to expand the functionality of existing orchard robots and of orchard robots currently under development in parallel research projects. The focus of the work is on two topics, namely the development of a sensing system and a gripper for picking pears and on a sensing system, robot control and end-effector(s) for robotic pruning of fruit trees and red currant bushes. On the longer term additional tasks such as automatic thinning, removing weeds and precision spraying will be targeted. This report specifically compiles three literature studies:

- Literature study pear detection methods (main author: Pieter Blok)
- Literature study robotic tree fruit harvesting (main author: Menno Sytsma)
- Literature study machine vision systems for robotic pruning (main author: Hitoshi Nishikawa)

2 Literature study pear detection methods

2.1 Introduction

To enable a robot to successfully harvest a pear, there is a need for a sensor and software system that can properly detect the pears. The most obvious sensor for this task is a Red-Green-Blue and Depth (RGB-D) camera, which can map the orchard environment with a combined colour and depth image. Typically, the colour image is used to detect the individual pears, whereas the depth image is used to estimate the distance between the camera and the pear, and also the size of the pear (needed to determine whether to harvest the pear or not). The most obvious software algorithm for the detection of the pears is a convolutional neural network (CNN). After successfully detecting the pear, a 3D post processing algorithm can be used to size the pear and to determine the distance between the camera and the pear.

For the robotic harvesting of pear, there are specific challenges, and some of these challenges influence the requirements of the pear detection algorithm:

1. It is desirable to also detect the peduncle of the pear, because pears need to be harvested with their peduncle attached (to prevent the pear from drying out and/or rot).
2. In the orchard there are high degrees of image occlusion, and this can cause the pear or peduncle to be occluded by other pears, plant organs (leaves, branches) or materials (wires, concrete poles).
3. Pears grow in clusters, indicating that there can an order in how to harvest the individual pears belonging to that cluster.

2.2 Literature studied

The following table gives and overview on the most relevant and recent publications on deep-learning based image analysis methods for fruit detection that have been studied in the scope of our research.

Table 1 *Overview of literature studied for fruit detection.*

Author and Year	Crop	Sensor used	CNN used	Detection principle	Performance metrics
(Yang et al., 2020)	Citrus	Kinect v2 (RGB-D)	Mask R-CNN and a branch segment merging algorithm	Simultaneously detect and measure citrus fruits and branches	The average precision of fruit and branch recognition are 88.15% and 96.27%. The average measurement error of fruits' transverse diameters, fruits' longitudinal diameters, and branch diameters are 2.52, 2.29, and 1.17 mm
(Nejati et al., 2020)	Kiwi	Basler ac1920-40uc USB 3.0 (RGB)	FCN-8s	Semantic segmentation of the calyx of the kiwifruit, branch and wire	F1 score of 0.82 on the typical lighting image set, but struggles with harsh lighting with an F1 score of 0.13. Utilising the prepossessing techniques the vision system under harsh lighting improves to an F1score 0.42. To address the fruit occlusion challenge, the overall approach was found to be capable of detecting 87.0% of non-occluded and 30.0% of occluded kiwifruit across all lighting conditions.
(Kang & Chen, 2019)	Apple	Kinect v2 (RGB-D)	DaSNet	Instance segmentation on the apples and then using a circle Hough transform	F1 score of 0.832 on the detection of apples
(Yin et al., 2021)	Grape	ZED camera	Mask R-CNN	Instance segmentation on the grapes and then using a RANSAC algorithm for grape cylinder model fitting, and the axis of the cylinder model was used to estimate the pose of the grape	The average precision, recall, and intersection over union (IOU) are 89.53, 95.33, and 82.00%, respectively on a test set of 210 images.
(Wagner et al., 2021)	Straw-berry	Simulation environment in Gazebo using the Intel RealSense	VGG16 network used for rotation estimation	Image analysis on single strawberry fruit clipped to a smaller image	Mean errors of as low as 11◦ could be achieved.
(Tu et al., 2020)	Passion fruit	Kinect v2	Faster R-CNN	Multiple scale faster region-based convolutional neural networks (MS-FRCNN) approach using the color and depth image	Compared with the faster R-CNN detector of RGB-D images, the recall, the precision and F1-score of MS-FRCNN method increased from 0.922 to 0.962, 0.850 to 0.931 and 0.885 to 0.946, respectively. Furthermore, the MS-FRCNN method effectively improves small passion fruit detection by achieving 0.909 of the F1 score.
(Lin et al., 2019)	Guava	Kinect V2	VGG-16-based FCN	A state-of-the-art fully convolutional network is first deployed to segment the RGB image to output a fruit and branch binary map. Based on the fruit binary map and RGB-D depth image, Euclidean clustering is then applied to group the point cloud into a set of individual fruits. Next, a multiple three-dimensional (3D) line-segments detection method is developed to reconstruct the segmented branches. Finally, the 3D pose of the fruit is estimated using its center position and nearest branch information.	Precision and recall of guava fruit detection were 0.983 and 0.948, respectively; the 3D pose error was 23.43° ± 14.18°

(Kang et al., 2020)	Apple	Intel Realsense D435	DaSNet	DaSNet takes raw input of RGB images from the RGBD camera to perform fruit detection and instance segmentation. The Pointnet grasping estimation takes the point cloud of each fruit as input, and predict the grasp pose for each fruit as output.	Pointnet grasping estimation achieves accurate grasp results on normal, noise, and outlier conditions, which are 0.91, 0.87, and 0.9, respectively.
(Yu et al., 2020)	Straw-berry	USB camera (type not specified)	R-YOLO	Fruit pose estimator called rotated YOLO (R-YOLO), which significantly improves the localization precision of the picking points	The test results of a set of 100 strawberry images showed that the proposed model's average recognition rate to be 94.43% and the recall rate to be 93.46%. Field test results showed that the harvesting success rate reached 84.35% in modified situations.
(Yu et al., 2019)	Straw-berry	Hand-held digital camera (type not specified)	Mask R-CNN	After generating mask images of ripe fruits from Mask R-CNN, a visual localization method for strawberry picking points was performed.	Fruit detection results of 100 test images showed that the average detection precision rate was 95.78%, the recall rate was 95.41% and the mean intersection over union (MIoU) rate for instance segmentation was 89.85%. The prediction results of 573 ripe fruit picking points showed that the average error was±1.2 mm.
(Weyler et al., 2021)	Sugar beet	Manta G-319 GigE	CenterNet and keypoint detection	CenterNet is used to detect the bounding boxes and then a keypoint detector is used to count the leaves inside the bounding box	The network shows superior performance for later growth stages of medium (97%) and large crops (67%) in contrast to Mask R-CNN (96% or 26%) allowing for a more accurate monitoring and determination of growth stages of individual plants.
(Sun et al., 2021)	Citrus	Mobile phone (Huawei and iPhone)	Faceboxes and multi-level feature fusion network (keypoint)	Object detection is done by FaceBoxes and then the cropped image is sent to a multi-level feature fusion network to detect keypoints	AP of 77.4% and an accuracy score of 84.7%

To this current date, there has not been any research conducted on the detection of pear with a CNN. Twelve other recent studies on fruit detection (apple, strawberry, kiwi) showed a diversity of used CNN's. They can be roughly divided into two groups: object detection and instance segmentation. With object detection, the most used algorithms were Faster R-CNN and YOLO. With instance segmentation, the most used algorithms were Mask R-CNN and DaSNet. Another study used a keypoint detection algorithm to identify relevant parts of the fruit and its peduncle.

Based on the literature review and the specific requirements for the detection of pear and its peduncle, we can identify four software methods to be used:

2.2.1 Instance segmentation + 3D vertex normal estimation

With this method, the RGB image is processed by an instance segmentation algorithm, for example Mask R-CNN. The instance segmentation algorithm results a polygon mask for each individual pear. This mask can be transferred to the depth image or the 3D point cloud constructed from the depth image. The result is a point cloud specific to the individual pear. Then, with a 3D post processing algorithm the normals of the vertices are determined, and the normal pointing upwards can indicate the location of the peduncle.

Pros:
- least annotation effort
- relatively easy software implementation
- the normal estimation is not influenced by occlusions (as long as the pear is detected)

Cons:
- with this method the peduncle is not directly detected, indicating that there is a risk that the method results inconsistent results
- other objects relevant for robotic harvest (like obstacles) cannot be detected

2.2.2 Extremely accurate instance segmentation

With this method, the RGB image is processed by an extremely accurate instance segmentation algorithm, for example Mask R-CNN PointRend. This algorithm allows the precise mask segmentation of both the pear and its peduncle.

Pros:
- relatively easy annotation
- most straightforward software implementation

Cons:
- when the peduncle is not visible it cannot be annotated and this might introduce model inconsistencies
- other objects relevant for robotic harvest (like obstacles) cannot be detected

2.2.3 Key-Point detection

With this method, the pear and its peduncle are detection by a combination of relevant keypoints. These keypoints are composed by a "skeleton" structure, indicating the hierarchy and the position of the points.

Pros:
- this method delivers the most "ready-to-use" output for the robot
- the size of the pear can be easily determined from the skeleton

Cons:
- when the peduncle is not visible, the human annotator has to reason about its location (and this can introduce inconsistencies)
- annotation takes significantly more time

- other objects relevant for robotic harvest (like obstacles) cannot be detected

2.2.4 Panoptic segmentation

- Panoptic segmentation is a combination of instance segmentation and semantic segmentation. With this method, the pears can be detected individually by the instance segmentation algorithm, whereas other relevant objects are pixel-segmented so that their location is known to the robot. The peduncles can be learned by the semantic segmentation.

Pros:
- with this method, other objects relevant for robotic harvest (like obstacles) can be detected
- the peduncles, which are typically thin and elongated objects, can be possibly learned better by the semantic segmentation compared to the standard Mask R-CNN segmentation

Cons:
- most time-consuming annotation process
- the panoptic segmentation algorithm in Detectron2 doesn’t allow yet training on custom datasets

Table 2 *Comparison different segmentation and detection methods.*

Score prognosis on a scale 1-5 (1 = bad, 5 = perfect)

Method	Peduncle detection	Ease of annotation	Ease of network training	Ease of practical implementation	Obstacle detection	Total score
Instance segmentation + 3D vertex normal estimation	2	5	5	2	1	15
Extremely accurate instance segmentation	4	4	5	5	1	19
Keypoint detection	5	2	3	5	1	16
Panoptic segmentation	5	1	1	4	5	16

2.3 Conclusion

It can be concluded that any of the four proposed methods can enable the detection of the pear and its peduncle, however taking everything into account, there is one method that stands out: the extremely accurate instance segmentation method (Mask R-CNN PointRend).

The next step is to investigate whether this method can achieve the best detection and pixel segmentation of the pears and the peduncles. If, for some reason, the method is less interesting than expected, then we will first investigate whether the keypoint detection method can be used as a better alternative. In case the detection of the obstacles has a higher urgency, then we will focus on panoptic segmentation. An important precondition is that the use of this algorithm depends on the possibility to train the algorithm on custom datasets.

3 Literature study robotic tree fruit harvesting

3.1 Introduction

Harvesting tasks are highly repetitive and very labour intensive. As they are also delicate tasks and the environment in which the harvesting of tree fruit needs to occur has a big variety a lot of research has been done to automate this without much commercially available harvesting solutions so far. The studies developing complete systems have focused largely on planning algorithms to reach fruits and grippers to ensure correct gripping without damaging the fruit (Davidson et al., 2020; Zhang et al., 2020). The review paper presented by Davidson for example links to several other studies presenting also the formal design processes used.

3.2 Harvesting robot functional requirements

Earlier studies have shown the following general functional requirements for a fruit harvesting robot (Davidson et al., 2020):

- Fruit detection
- Approach fruit
- Detach fruit
- Guide harvested product

The fruit detection system is covered in another document within this project. The other requirements result in research fields on path planning, gripper design and fruit manipulation patterns.

3.3 Methods for approaching of fruit/target

Most robotic harvesting systems make use of robotic arms or a linear system that gives the end-effector at least 6 degrees of freedom (Bac et al., 2014; Roldán et al., 2018). For the path planning most studies rely on rapid exploring random trees (RRT) (Nemlekar et al., 2021), which they combine with a method to check whether the robot is in collision. For this the Flexible Collision Library (Pan et al., 2012) is used, but also collision detectors based on geometric primitives are used (Paulin et al., 2015) as they are computationally cheaper than complex meshes. Tools used to implement the path planning algorithms are MATLAB (Masood & Haghshenas-Jaryani, 2021) and ROS. For the last one more specifically the RRT-Connect implementation from the Open Motion Planning Library was mentioned (Botterill et al., 2017). Other than solving the path planning problem studies also focused on the reachability of fruits and how this can be affected by the size of a gripper (Arikapudi & Vougioukas, 2021).

3.4 Gripper requirements

For the gripper itself many different solutions have been proposed (Navas et al., 2021; Shintake et al., 2018; Zhang et al., 2020). Presented prototypes vary from vacuum tube end-effectors (Curt Salisbury, 2014), soft passive fingers (Yu et al., 2021), bio-inspired grippers, (Pi et al., 2021) and rigid-finger systems (Davidson et al., 2015). The review papers are all trying to structure the broad range of proposed solutions. Different prototypes are evaluated based on different criteria by Navas (Navas et al., 2021), which can be translated in the following requirements for a pear harvesting gripper:

- Object size: Gripper should be able to pick various pear sizes

- Gripper size
- Lifting ratio
- Power consumption
- Scalability (ease of manufacturing/ modularity of technology used)
- Controllability
- Response time
- Surface conditions
- Degree of skill to work in unstructured environments
- Mechanical compliance
- Lifetime
- Technology readiness level

End-effectors aiming to grasp the fruit are aiming to perform a power grasp for spherical objects. A method to classify human grasping methods which widely used is presented by Cutcosky (Cutkosky, 1989). In the performance evaluation of grippers most studies limit themselves to the force that can be applied on the grasped objects without damaging them, what weight of objects can be lifted and how different end-effector types can affect this. The interactions are measured using pressure sensors on the fingers (Davidson et al., 2015; Pi et al., 2021; Teeple et al., 2020). Practical examples of how grippers respond to non-isolated objects are hardly shown.

3.5 Manipulation patterns

No literature was found on manipulation patterns specifically for pears, but methods to get more knowledge of picking patterns are available for other fruits. Classification method for harvesting movements and basic picking patterns are mostly decomposing the complete movement in smaller simple steps (Navas et al., 2021). Another topic of interest is the minimal force required to detach a fruit from the plant or tree. For Tomatoes and apples studies have been performed to determine an optimal deflection angle and applied force using pressure sensors and force gauges applied on top of an end-effector or human hand (Davidson et al., 2016; Li et al., 2016; Xie et al., 2021). In the use case of cherry tomatoes a similarity with pears can be found. For cherry tomatoes it is relevant that the fruit is separated from the plant at the calyx, this study therefore also aimed at picking patterns with focus on this separation. The sensors used are a GripTM Tekscan tactile sensor, flexible force sensors from Tekscan, HP-200 HANDPI force sensor, an Imu. Data in these studies was processed using Matlab’s Statistics and machine learning toolbox.

3.6 Questions raised from this literature study

Reachability has been studied, but methods to improve this have not. Many grippers show promising results in their grasping properties, but they do not seem to break through. Grippers are working properly under ideal circumstances, but in a real setup circumstances are far from ideal. Is it possible to do something about that? What is needed for the gripper to properly grip the target object or make it more reachable?

For pears it is still not exactly known what is needed to harvest them. Conversations with pear growers have indicated that keeping the peduncle intact is a much more critical element than is the case for harvesting apples. Is it possible to do a similar approach as with the tomato-case to determine the optimal detachment point that does not harm the peduncle?

Research has been performed on what forces are needed to detach an apple, but no examples have been found where it is also converted to controlling a robot to apply the same delicate movements as are performed by harvesters.

4 Literature study machine vision systems for robotic pruning

4.1 Introduction

Pruning of fruit bearing trees and bushes is done by selective branch removal for the purpose of work management, growth control, keeping quality and uniformity of fruit, and pest removal. Securing skilled workers and their high labor costs have been major issues, and pruning automation is an important research topic. The main research fields of pruning automation can be broadly divided into Machine Vision System and Pruning Robot. The Machine Vision System aims to model the shape of a tree using a sensor system and a data processing algorithm, to recognize trunks or branches and to specify the points to be cut. Pruning robots aim to cut specific points using end effectors attached to the robot arm, including path planning including Obstacle Avoidance. In this chapter, we focus on Machin Vision Systems.

4.2 Machine Vision Systems

The Machine Vision System can be divided into “Sensing System” using cameras, “Data Pre-Processing” including noise reduction, down sampling or 3D reconstruction, and “Recognition Functions” including branch segmentation or key-point detection. It is usually a pipeline process, and the previous process affects the performance of subsequent processes. Since the difficulty varies depending on the conditions and target crops, it is necessary to pay attention to the performance comparison shown in the literature.

4.2.1 Sensing System

Two-dimensional based techniques only from RGB images have been studied so far, but they have not been sufficient to locate branches in 3D space (Gao & Lu, 2006). In the current research, the use of stereo cameras and RGB-D cameras is the most mainstream. An example of configuring a system with only RGB images is seen in (Botterill et al., 2017). In this example, they proposed a platform that reconstructs a 3D model of a tree by triangulation from images taken by three RGB cameras shaded by a straddling platform, and identifies pruning points by a geometric method. Simple RGB-D cameras such as ZED2 Stereolabs (He et al., 2021) or Intel RealSense devices (Intel, 2022) are susceptible to ambient light, which may be a limitation for outdoor orchard use where imaging in natural and changing light is required. On the other hand, depth cameras and also ToF (Time-of-Flight) cameras have been reported to have good performance, and considering the price, RealSense (Chen et al., 2020) and Kinect v2 (Majeed et al., 2018) can be said to be feasible candidates at present. Multiple RGB-D cameras can be used in combination to improve the accuracy of the point cloud (Ma et al., 2021). Cuevas Velásquez (Cuevas Velásquez et al., 2020) used a single stereo RGB camera mounted on a robotic arm to image a rose bush from multiple viewpoints to minimize occlusions. LiDAR is also often used for this purpose and can be said to be one of the feasible sensors because it can obtain an accurate point cloud (Westling et al., 2021). However, LIDAR sensors often only produce geometric point cloud data without colour information. Still, colour information might be essential for determine the cutting locations. Laser scanners (Bohn Reckziegel et al., 2022; Boogaard et al., 2021; Dutagaci et al., 2020) can obtain a very accurate point cloud, but it is difficult to scan in real time, so it is only used for research purposes of phenotyping. The price is also the reason of its difficulty for practical use.

As a different sensor, research on branch detection for 1 and 2 years using a hyperspectral camera is interesting (Khanal et al., 2018). In addition to the RGB visible wavelength, it was shown that there was a particularly large difference in the water absorption wavelength. However, this method has problems in selecting a camera that can be used outdoors and setting measurement conditions such as separation from the background.

4.2.2 Data pre-processing

One of the main purposes of data pre-processing is to create a high quality point cloud to improve the accuracy of subsequent key-point detection. In the case of a sensing system using an RGB camera, the photogrammetry method is generally used, but it lacks real-time performance and is mainly used for phenotyping applications (Gené-Mola et al., 2021; Straub et al., 2021; Zine-El-Abidine et al., 2020). In the case of a method of applying deep learning to a data down sampling 3D point cloud in order to realize high-speed processing, the feasible data size is often about 4k-40k. In this case, Voxel down sampling is performed as needed before Key-point Detection (Ge et al., 2021). In addition to down sampling, consider data division as necessary (Boogaard et al., 2021; Turgut et al., 2022). In the case of a system that aims to improve accuracy by using multiple depth cameras, i.e. multiple point clouds, an alignment method called point cloud registration is usually used. Since this algorithm often takes a long time to process, a method of performing rough alignment by global registration before detailed alignment by ICP (Iterative Closest Point) or alignment method with feature point extraction called FPFH (Fast Point Feature Histograms) is used (Ma et al., 2021). Data annotation is also included in the data pre-process in researches using deep learning. In the researches of 2D images, many researches are related to the efficiency of annotation such as semi-supervised learning, weak supervised learning, or active learning have been done. On the other hand, in research dealing with 3D point clouds, manual annotation by OSS (Open Source Software) such as Cloud Compare is currently the mainstream, and human cost is high (Bohn Reckziegel et al., 2022; Boogaard et al., 2021; Ma et al., 2021). In case of instant segmentation of 2D images, OSS (Open Source Software) applications such as VGG image annotator (Lin et al., 2021), COCO annotator (Fernandes et al., 2021) etc. are used.

4.2.3 Key-point Detection

Key-points are interest points in images. They are points in the image that define what is interesting or what stand out in the image and are often used for analyzing and interpreting the image. Concerning key-point detection methods it can be said that the 2D-based methods or methods of combining the 2D-based methods with the 3D-based methods ae one of the feasible methods from the viewpoint of achieving both recognition speed and accuracy rate (Botterill et al., 2017; Chen et al., 2020; Fernandes et al., 2021; Lin et al., 2021; Majeed et al., 2018). As a method for directly handling the point cloud, a method for analyzing the structure by a geometric method such as DBSCAN or ICP optimization is often used (Lin et al., 2021; Straub et al., 2021; Zine-El-Abidine et al., 2020). Methods such as TreeQSM (InverseTampere, 2020) and GraphTreeLS (Westling et al., 2021), which analyze the structure of a tree by a geometric method in combination with knowledge of the tree structure, have also been proposed, including prerequisite knowledge about the structure of the tree to adapt to the target tree. Detailed tuning of the parameters is required. In the case of Bohn Reckziegel (Bohn Reckziegel et al., 2022), the model structure of the tree was analyzed using TreeQSM, and the simulation of the pruning method and its effect on the light captured in the canopy were performed. In recent years, methods directly applying deep learning to 3D point cloud have also been studied. (Turgut et al., 2022) compared the deep learning models available in the point cloud and showed that PointNet or its improved version, PointNet++, provides good accuracy. PointNett ++ is also used in (Boogaard et al., 2021) because of its high segmentation accuracy. In (Ma et al., 2021), SPGNet model was used to realize 3D segmentation of Trunk and Branch from the point cloud. As an example of another 2D-based method, the simulation method utilizing reinforcement learning with 2D images is interesting (You et al., 2021). Through 2D-based reinforcement learning in the simulation domain, pruning behavior of robot arm and end effector can be learned without directly dealing with 3D models. This has a big advantage in that it does not need for real-world iterations. On the other hand, it is considered that human cost is required to create a simulation model that is faithful to the actual model.

Table 3 *Overview publications on detection methods for robotic tree pruning.*

Reference	crop	Sensors	application	2D/3D	real time	methodology / tools
(Fernandes et al., 2021)	grapevine	RGB camera	pruning point suggestion	2D	yes	COCO anotator Detectron2
(Gené-Mola et al., 2021)	apple		fruit size estimation	2D+3D	no	SfM and MVS (MetaShape) CloudCompare Mask-RCNN
(Zine-El-Abidine et al., 2020)	apple		count/locate fruits	3D	no	VisualSFM Color Checker Passport Photo 2 Iterative Closest Point (ICP)
(Straub et al., 2021)	apple		phenotyping (tree modeling)	3D	no	Metashape RANSAC / RandomForest / k-means Cloud Compare
(Botterill et al., 2017)	grapevine	RGB camera (multi)	pruning	2D+3D	yes	RBF-SVM / Hough transform Stocastic Image Grammar method LevenbergMarquardt optimization
(He et al., 2021)	straw-berry	RGB-D camera ZED2 stereo camera	maturity estimation	2D	yes	YOLOv4 / Alexnet labelimg_v1.8.0
(Chen et al., 2020)	apple	RGB-D camera Intel RealSense D435	pruning (tree modeling)	2D	yes	U-Net / DeepLabv3/ GAN (Pix2Pix) Adobe Photoshop
(You et al., 2021)	sweet cherry		pruning	2D	yes	reinforcement learning by DNN GAN / PyBullet / ilastik
(Ge et al., 2021)	apple		yield estimation	3D	no	VisualSFM MeshLab / FPFH SVM / PCL / MATLAB
(Majeed et al., 2018)	apple	RGB-D camera Kinect V2	pruning (segmentation)	2D	no	SegNet (fine-tuned by ImageNet) MATLAB libraries
(Lin et al., 2021)	guava		harvesting		yes	MASK-RCNN / RANSAC / PCA OpenCV / VGG image annotator
(Ma et al., 2021)	jujube	RGB-D camera Kinect V2 (multi)	pruning (segmentation)	3D	no	SPGNet Laplacian algorithm FPFH / ICP Cloud Compare
(Westling et al., 2021)	abocado mango	LiDAR	prunig (pruning strategy)	3D	no	SimTreeLS GraphTreeLS
(Bohn Reckziegel et al., 2022)	cherry	Laser scanner	pruning	3D	no	TreeQSM LaserControl CloudCompare
(Boogaard et al., 2021)	cucumber	Laser sensor (multi) Spectral sensor	phenotyping	3D	no	PointNet++ CloudCompare
(Khanal et al., 2018)	red raspberry	Hyper spectral sensor	pruning	2D	no	Linear SVM (from MATLAB Toolbox) PCA / Random Frog
(Dutagaci et al., 2020)	rose bush	X-ray sensor	phenotyping	3D	no	-
(Turgut et al., 2022)	rose bush		phenotyping (segmentation)	3D	no	PointNet / PointNet++ / DgCNN PointCNN / SehllNet / RIConv

References

Arikapudi, R., & Vougioukas, S. G. (2021). Robotic Tree-Fruit Harvesting With Telescoping Arms: A Study of Linear Fruit Reachability Under Geometric Constraints. *IEEE Access, 9*, 17114-17126.

Bac, C. W., van Henten, E. J., Hemming, J., & Edan, Y. (2014). Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics, 31*(6), 888-911.

Bohn Reckziegel, R., Sheppard, J. P., Kahle, H.-P., Larysch, E., Spiecker, H., Seifert, T., & Morhart, C. (2022). Virtual pruning of 3D trees as a tool for managing shading effects in agroforestry systems. *Agroforestry Systems, 96*(1), 89-104. doi:10.1007/s10457-021-00697-5 Retrieved from https://doi.org/10.1007/s10457-021-00697-5.

Boogaard, F. P., van Henten, E. J., & Kootstra, G. (2021). Boosting plant-part segmentation of cucumber plants by enriching incomplete 3D point clouds with spectral data. *Biosystems Engineering, 211*, 167-182. doi:https://doi.org/10.1016/j.biosystemseng.2021.09.004 Retrieved from https://www.sciencedirect.com/science/article/pii/S1537511021002233.

Botterill, T., Paulin, S., Green, R., Williams, S., Lin, J., Saxton, V., . . . Corbett-Davies, S. (2017). A robot system for pruning grape vines. *Journal of Field Robotics, 34*(6), 1100-1122.

Chen, Z., Ting, D., Newbury, R., & Chen, C. (2020). Semantic Segmentation for Partially Occluded Apple Trees Based on Deep Learning. doi:10.48550/ARXIV.2010.06879 Retrieved from https://arxiv.org/abs/2010.06879.

Cuevas Velásquez, H., Gallego, A. J., Tylecek, R., Hemming, J., Tuijl, B., Mencarelli, A., & Fisher, R. (2020). *Real-time Stereo Visual Servoing for Rose Pruning with Robotic Arm.* Paper presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA).doi:10.1109/ICRA40945.2020.9197272

Curt Salisbury, P. (2014). *Evaluation of vacuum as a picking mechanism*. Retrieved from Menlo Park, CA 94025

Cutkosky, M. R. (1989). On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on robotics and automation, 5*(3), 269-279.

Davidson, J., Bhusal, S., Mo, C., Karkee, M., & Zhang, Q. (2020). Robotic Manipulation for Specialty Crop Harvesting: A Review of Manipulator and End-Effector Technologies. *Global Journal of Agricultural and Allied Sciences, 2*(1), 25-41.

Davidson, J., Silwal, A., Karkee, M., Mo, C., & Zhang, Q. (2016). Hand-picking dynamic analysis for undersensed robotic apple harvesting. *Transactions of the ASABE, 59*(4), 745-758.

Davidson, J. R., Mo, C., Silwal, A., Karkee, M., Li, J., Xiao, K., . . . Lewis, K. (2015). *Human-machine collaboration for the robotic harvesting of fresh market apples.* Paper presented at the Proceedings of IEEE ICRA workshop on robotics in agriculture

Dutagaci, H., Rasti, P., Galopin, G., & Rousseau, D. (2020). ROSE-X: an annotated data set for evaluation of 3D plant organ segmentation methods. *Plant Methods, 4*(16), 28. doi:10.1186/s13007-020-00573-w.

Fernandes, M., Scaldaferri, A., Fiameni, G., Teng, T., Gatti, M., Poni, S., . . . Chen, F. (2021). Grapevine Winter Pruning Automation: On Potential Pruning Points Detection through 2D Plant Modeling using Grapevine Segmentation. doi:10.48550/ARXIV.2106.04208 Retrieved from https://arxiv.org/abs/2106.04208.

Gao, M., & Lu, T. (2006, 25-28 June 2006). *Image Processing and Analysis for Autonomous Grapevine Pruning.* Paper presented at the 2006 International Conference on Mechatronics and Automation.doi:10.1109/ICMA.2006.257748

Ge, L., Zou, K., Zhou, H., Yu, X., Tan, Y., Zhang, C., & Li, W. (2021). Three dimensional apple tree organs classification and yield estimation algorithm based on multi-features fusion and support vector machine. *Information Processing in Agriculture*. doi:https://doi.org/10.1016/j.inpa.2021.04.011 Retrieved from https://www.sciencedirect.com/science/article/pii/S2214317321000354.

Gené-Mola, J., Sanz-Cortiella, R., Rosell-Polo, J. R., Escolà, A., & Gregorio, E. (2021). In-field apple size estimation using photogrammetry-derived 3D point clouds: Comparison of 4 different methods considering fruit occlusions. *Computers and Electronics in Agriculture, 188*, 106343. doi:https://doi.org/10.1016/j.compag.2021.106343 Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169921003604.

He, Z., Karkee, M., & Upadhyaya, P. (2021). *Detection of strawberries with varying maturity levels for robotic harvesting using YOLOv4*. Paper presented at the 2021 ASABE Annual International Meeting. doi:10.13031/aim.202100051

Intel. (2022). Intel RealSense camera overview. Retrieved from https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html accessed

InverseTampere. (2020). TreeQSM. Retrieved from https://github.com/InverseTampere/TreeQSM, accessed

Kang, H., & Chen, C. (2019). Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards. *Sensors, 19*(20), 4599. doi:10.3390/s19204599 Retrieved from https://www.mdpi.com/1424-8220/19/20/4599.

Kang, H., Zhou, H., Wang, X., & Chen, C. (2020). Real-Time Fruit Recognition and Grasping Estimation for Robotic Apple Harvesting. *Sensors, 20*(19), 5670. Retrieved from https://www.mdpi.com/1424-8220/20/19/5670.

Khanal, K., Bhusal, S., Karkee, M., & Zhang, Q. (2018). Distinguishing One Year and Two Year Old Canes of Red Raspberry Plant using Spectral Reflectance. *IFAC-PapersOnLine, 51*(17), 39-44. doi:https://doi.org/10.1016/j.ifacol.2018.08.058 Retrieved from https://www.sciencedirect.com/science/article/pii/S2405896318311741.

Li, J., Karkee, M., Zhang, Q., Xiao, K., & Feng, T. (2016). Characterizing apple picking patterns for robotic harvesting. *Computers and Electronics in Agriculture, 127*, 633-640.

Lin, G., Tang, Y., Zou, X., & Wang, C. (2021). Three-dimensional reconstruction of guava fruits and branches using instance segmentation and geometry analysis. *Computers and Electronics in Agriculture, 184*, 106107. doi:10.1016/j.compag.2021.106107.

Lin, G., Tang, Y., Zou, X., Xiong, J., & Li, J. (2019). Guava Detection and Pose Estimation Using a Low-Cost RGB-D Sensor in the Field. *Sensors, 19*(2). doi:10.3390/s19020428.

Ma, B., Du, J., Wang, L., Jiang, H., & Zhou, M. (2021). Automatic branch detection of jujube trees based on 3D reconstruction for dormant pruning using the deep learning-based method. *Comput. Electron. Agric., 190*(C), 12. doi:10.1016/j.compag.2021.106484 Retrieved from https://doi.org/10.1016/j.compag.2021.106484.

Majeed, Y., Zhang, J., Zhang, X., Fu, L., Karkee, M., Zhang, Q., & Whiting, M. D. (2018). Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation. *IFAC-PapersOnLine, 51*(17), 75-80. doi:https://doi.org/10.1016/j.ifacol.2018.08.064 Retrieved from https://www.sciencedirect.com/science/article/pii/S2405896318311807.

Masood, M. U., & Haghshenas-Jaryani, M. (2021). A Study on the feasibility of robotic harvesting for chile pepper. *Robotics, 10*(3), 94.

Navas, E., Fernández, R., Sepúlveda, D., Armada, M., & Gonzalez-de-Santos, P. (2021). Soft Grippers for Automatic Crop Harvesting: A Review. *Sensors, 21*(8), 2689.

Nejati, M., Penhall, N., Williams, H., Bell, J., Lim, J., Ahn, H. S., & MacDonald, B. A. (2020). Kiwifruit detection in challenging conditions. *CoRR, abs/2006.11729*. Retrieved from https://arxiv.org/abs/2006.11729.

Nemlekar, H., Liu, Z., Kothawade, S., Niyaz, S., Raghavan, B., & Nikolaidis, S. (2021). Robotic Lime Picking by Considering Leaves as Permeable Obstacles. *arXiv preprint arXiv:2108.13889*.

Pan, J., Chitta, S., & Manocha, D. (2012). *FCL: A general purpose library for collision and proximity queries.* Paper presented at the 2012 IEEE International Conference on Robotics and Automation

Paulin, S., Botterill, T., Chen, X., & Green, R. (2015). *A specialised collision detector for grape vines.* Paper presented at the Proceedings of the Australasian Conference on Robotics and Automation

Pi, J., Liu, J., Zhou, K., & Qian, M. (2021). An Octopus-Inspired Bionic Flexible Gripper for Apple Grasping. *Agriculture, 11*(10), 1014.

Roldán, J. J., del Cerro, J., Garzón-Ramos, D., Garcia-Aunon, P., Garzón, M., de León, J., & Barrientos, A. (2018). Robots in agriculture: State of art and practical experiences. *Service robots*, 67-90.

Shintake, J., Cacucciolo, V., Floreano, D., & Shea, H. (2018). Soft robotic grippers. *Advanced Materials, 30*(29), 1707035.

Straub, J., Reiser, D., & Griepentrog, H. W. (2021). *Approach for modeling single branches of meadow orchard trees with 3D point clouds*. Paper presented at the Precision agriculture '21. https://doi.org/10.3920%2F978-90-8686-916-9_88. doi:10.3920/978-90-8686-916-9_88

Sun, Q., Chai, X., Zeng, Z., Zhou, G., & Sun, T. (2021). Multi-level feature fusion for fruit bearing branch keypoint detection. *Computers and Electronics in Agriculture, 191*, 106479. doi:https://doi.org/10.1016/j.compag.2021.106479 Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169921004968.

Teeple, C. B., Koutros, T. N., Graule, M. A., & Wood, R. J. (2020). Multi-segment soft robotic fingers enable robust precision grasping. *The International Journal of Robotics Research, 39*(14), 1647-1667.

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., . . . Xue, Y. (2020). Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. *Precision Agriculture, 21*(5), 1072-1091. doi:10.1007/s11119-020-09709-3 Retrieved from https://doi.org/10.1007/s11119-020-09709-3.

Turgut, K., Dutagaci, H., Galopin, G., & Rousseau, D. (2022). Segmentation of structural parts of rosebush plants with 3D point-based deep learning methods. *Plant Methods, 18*(1), 20. doi:10.1186/s13007-022-00857-3 Retrieved from https://doi.org/10.1186/s13007-022-00857-3.

Wagner, N., Kirk, R., Hanheide, M., & Cielniak, G. (2021, 30 May-5 June 2021). *Efficient and Robust Orientation Estimation of Strawberries for Fruit Picking Applications.* Paper presented at the 2021 IEEE International Conference on Robotics and Automation (ICRA).doi:10.1109/ICRA48506.2021.9561848

Westling, F., Underwood, J., & Bryson, M. (2021). A procedure for automated tree pruning suggestion using LiDAR scans of fruit trees. doi:10.48550/ARXIV.2102.03700 Retrieved from https://arxiv.org/abs/2102.03700.

Weyler, J., Milioto, A., Falck, T., Behley, J., & Stachniss, C. (2021). Joint Plant Instance Detection and Leaf Count Estimation for In-Field Plant Phenotyping. *IEEE Robotics and Automation Letters, 6*(2), 3599-3606. doi:10.1109/LRA.2021.3060712.

Xie, H., Kong, D., Shan, J., & Xu, F. (2021). Study the Parametric Effect of Pulling Pattern on Cherry Tomato Harvesting Using RSM-BBD Techniques. *Agriculture, 11*(9), 815.

Yang, C. H., Xiong, L. Y., Wang, Z., Wang, Y., Shi, G., Kuremot, T., . . . Yang, Y. (2020). Integrated detection of citrus fruits and branches using a convolutional neural network. *Computers and Electronics in Agriculture, 174*, 105469. doi:https://doi.org/10.1016/j.compag.2020.105469 Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169920300211.

Yin, W., Wen, H., Ning, Z., Ye, J., Dong, Z., & Luo, L. (2021). Fruit Detection and Pose Estimation for Grape Cluster–Harvesting Robot Using Binocular Imagery Based on Deep Neural Networks. *Frontiers in Robotics and AI, 8*. doi:10.3389/frobt.2021.626989 Retrieved from https://www.frontiersin.org/article/10.3389/frobt.2021.626989.

You, A., Kolano, H., Parayil, N., Grimm, C., & Davidson, J. (2021). Precision fruit tree pruning using a learned hybrid vision/interaction controller. doi:10.48550/arXiv.2109.13162.

Yu, X., Fan, Z., Wang, X., Wan, H., Wang, P., Zeng, X., & Jia, F. (2021). A lab-customized autonomous humanoid apple harvesting robot. *Computers & Electrical Engineering, 96*, 107459.

Yu, Y., Zhang, K., Liu, H., Yang, L., & Zhang, D. (2020). Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot. *IEEE Access, 8*, 116556-116568. doi:10.1109/ACCESS.2020.3003034.

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture, 163*, 104846. doi:https://doi.org/10.1016/j.compag.2019.06.001 Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169919301103.

Zhang, B., Xie, Y., Zhou, J., Wang, K., & Zhang, Z. (2020). State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review. *Computers and Electronics in Agriculture, 177*, 105694.

Zine-El-Abidine, M., Dutagaci, H., Galopin, G., & Rousseau, D. (2020). Assigning Apples to Individual Trees in Dense Orchards using 3D Color Point Clouds. doi:10.48550/ARXIV.2012.13721 Retrieved from https://arxiv.org/abs/2012.13721.

To explore
the potential
of nature to
improve the
quality of life

Wageningen University & Research
Corresponding address for this report:
P.O. Box 16
6700 AA Wageningen
The Netherlands
T +31 (0)317 48 07 00
www.wur.eu/plant-research

Report WPR-987

De missie van Wageningen University & Research is 'To explore the potential of nature to improve the quality of life'. Binnen Wageningen University & Research bundelen Wageningen University en gespecialiseerde onderzoeksinstituten van Stichting Wageningen Research hun krachten om bij te dragen aan de oplossing van belangrijke vragen in het domein van gezonde voeding en leefomgeving. Met ongeveer 30 vestigingen, 7.200 medewerkers (6.400 fte) en 13.200 studenten en ruim 150.000 Leven Lang Leren-deelnemers behoort Wageningen University & Research wereldwijd tot de aansprekende kennisinstellingen binnen haar domein. De integrale benadering van de vraagstukken en de samenwerking tussen verschillende disciplines vormen het hart van de unieke Wageningen aanpak.