Dynamic Sensing and Task Planning for an Autonomous Harvesting Robot

Thesis submitted in partial fulfillment
of the requirements for the degree of
"DOCTOR OF PHILOSOPHY"
by

Polina Kurtser

Submitted to the Senate of
Ben-Gurion University of the Negev

October 10, 2018
Beer-Sheva

Dynamic Sensing and Task Planning for an Autonomous Harvesting Robot

Thesis submitted in partial fulfillment
of the requirements for the degree of
"DOCTOR OF PHILOSOPHY"
by

Polina Kurtser

Submitted to the Senate of
Ben-Gurion University of the Negev

Approved by the advisor 
Approved by the Dean of the Kreitman School of Advanced Graduate Studies

October 10, 2018
Beer-Sheva

This work was carried out under the supervision of Prof. Yael Edan in the Department of Industrial Engineering and Management, Faculty of Engineering Sciences

I, Polina Kurtser, whose signature appears below, hereby declare that:

- ✓ I have written this thesis by myself, except for the help and guidance offered by my thesis advisors.
- ✓ The scientific materials included in this thesis are products of my own research, culled from the period during which I was a research student.

Date: October 10, 2018 Student's name: Polina Kurtser Signature: P.K.

To my mom, without you I would never be where I am, I would never be who I am.
To my partner, I wouldn't have met you without it, it would be much less fun without you.

> Если я человека люблю, я хочу, чтоб ему от меня стало лучше - хотя бы пришитая пуговица. От пришитой пуговицы - до всей моей души.
>
> — Марины И. Цветаевой
> (Russian poet)

Acknowledgements

First and foremost my deepest gratitude to my great advisor, Professor Yael Edan. Thank you for pushing me hard when needed, and stepping back to let me grow. Thank for being a constant source of knowledge, energy and experience. For being most professional for one moment and sharing a bunk bed in the middle of nowhere in the far-far north the next. For all the tiring field experiments done together, and the beers after them. You are the example of honesty and modesty I wish to preserve.
Thank you to the whole SWEEPER team. Your professionalism, work ethics and true cooperation abilities have proved how a real working product can be delivered against all chances. In particular Boaz Arad (a.k.a MacGyver) who saved so many experiments with a smile and grace! One can only wish for a colleague like you.
I'm grateful to my PhD committee - Professor Yisrael Paramet and Professor Dvir Shabtai. Thank you for your advise, guidance and great sense of humor. Special thank you to Professor Emeritus Nava Pliskin for great mentoring and writing advices.
Big thank you to my lab mates and colleagues at Ben Gurion University - Faina, Michael, Samuel, Vesna, Ben, Dor, Shahar, Omri and Shanee. You made this period so much more interesting. Thank you for being there for me when it was tough and happy for me when it was a bit easier.
And finally to my closest ones. Ola, thank you for listening supporting and pushing me to not compromise. Mom, thank you for baring with my ramblings, coping with my never-ending traveling, forgiving and accepting. I love you both very much.

This research was supported by the European Community's Horizon 2020 Framework Programme under the SWEEPER project (grant agreement no.644313). Partial support was also provided by the ABC Robotics Initiative funded by Helmsley Charitable Trust and by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, both at Ben-Gurion University of the Negev.

Table of Contents

1 Introduction **1**
- 1.1 Problem formulation . . . 2
- 1.2 Research objectives . . . 6
- 1.3 Innovation and contribution . . . 7
- 1.4 Thesis Structure . . . 9

2 Background **11**
- 2.1 Autonomous robotic harvesters . . . 11
- 2.2 Fruit detection from single and multiple viewpoints . . . 12
- 2.3 Harvesting cycle times . . . 13
- 2.4 Traveling salesman problem in agriculture . . . 14
- 2.5 Data acquisition . . . 15

3 Research methods **17**
- 3.1 Overview . . . 17
- 3.2 Data acquisition . . . 19
- 3.3 Acquired datasets . . . 21

4 Statistical models for fruit detectability: spatial and temporal analyses **23**
- 4.1 Introduction . . . 24
- 4.2 Background . . . 25
- 4.3 Statistical model development methodology . . . 28
- 4.4 Detectability analysis for a case study of sweet peppers . . . 38
- 4.5 Conclusions . . . 46
- 4.6 Research questions answered . . . 49

5 The use of dynamic sensing strategies to improve detection **51**
- 5.1 Introduction . . . 51
- 5.2 Dynamic sensing algorithm . . . 53
- 5.3 Evaluation methods . . . 60
- 5.4 Results . . . 64
- 5.5 Conclusions . . . 65
- 5.6 Research questions answered . . . 66

6 Planning the sequence of tasks **69**
6.1 Introduction 69
6.2 Formalization and algorithm 71
6.3 Methods 75
6.4 Results 80
6.5 Conclusions 83
6.6 Research questions answered 83

7 Evaluation of approach strategies for harvesting robots **85**
7.1 Introduction 86
7.2 Approach strategies 88
7.3 Experimental methods 90
7.4 Results 96
7.5 Discussion 102
7.6 Conclusions 105
7.7 Research questions answered 106

8 Conclusions & Recommendations **107**
8.1 Main results 107
8.2 Recommendations 107
8.3 Limitations 108
8.4 Future work 109
8.5 Application to other crops 110

9 Bibliography **111**

10 Appendixes **121**
10.1 List of software and databases 121
10.2 SWEEPER public deliverable 123
10.3 Approach direction - conference proceeding 133

List of Figures

1.1 Autonomous pepper harvester sensing-planning-acting paradigm breakdown and the defined challenges. . . . 5
3.1 Robotic arm, sensor and lighting configuration. . . . 20
3.2 The 14 viewpoints from which image acquisition is performed. . . . 20
3.3 Data acquisition setup. Left: Isle in a sweet pepper commercial greenhouse. Right: Prof. Yael Edan, Dr. Boaz Arad and me during data acquisition. . . . 21
3.4 Two example images taken from the same distance from two viewpoints. Left: an image taken from the right. Right: an image taken from the left. . . . 21
4.1 Flowchart of overall statistical modeling of detectability. . . . 29
4.2 Definition of a viewpoint by distance and two angles – tilt and azimuth. Left: the definitions; Middle and Right: estimations by defining the distance to be the distance to the isle and the azimuth and tilt angle to be θ_a and θ_t accordingly as seen from a top and side views of the greenhouse lane. . . . 31
4.3 Peppers to be harvested from each side of the stems are those between the robot arm and the stems. Given a stem oriented south-north the peppers to be harvested from lane 1 are facing west and the ones to be harvested in lane 2 are facing east. . . . 31
4.4 Images taken in different seasonal conditions. Left: beginning of season, 3 days since last harvesting; Middle: end of season, harvesting performed 2 days prior acquisition; Right: challenging lighting conditions when images taken against the light. . . . 32
4.5 Acquisition protocol. . . . 33
4.6 Viewpoints covering area with no occlusion present. . . . 33
4.7 Example of a set splitting into training and test sets. . . . 35
4.8 Example of a tagged image and the classification of the marked peppers into ROIs of 3 types: fully visible, partially occluded, and truncated. . . . 36
4.9 Labeling process - window based manual tagging. . . . 40
4.10 Same image taken with artificial illumination (left) and natural overexposed light (right). . . . 40
4.11 Distribution of RA_l in DB#2 (Left) and DB#3 (Right). . . . 41
4.12 Detectability results, single viewpoint. . . . 44
5.1 Examples of visibility variability in different crop types. Left: Cherry tomatoes, high visibility. Right: Grapes, mid/low visibility. . . . 54
5.2 Image taken in a greenhouse including 3 peppers, frontal leaves, side leaves, and additional occluding items. . . . 54
5.3 Results of k-means classification on a transformed RGB image. Groups 1-3 represent the classification into peppers, occlusion and background. . . . 55
5.4 Three clusters classified by the k-means procedure. Left: occlusion, middle: peppers, right: background. . . . 55
5.5 Resulted filtered pepper mask. $N_{RP} = 1$. . . . 56

5.6 Noise reduction procedure on occlusion cluster. 56
5.7 Threshold by area and aspect-ratio. 56
5.8 Change in yaw angle in highly occluded scenes. Yaw angles left: $45°$, right: $135°$. 57
5.9 Same scene with 3 viewpoints decreasing in distance between camera and stems from left to right. Left: 19cm, middle: 10cm, right: 7cm. 57
5.10 Next viewpoint decision tree. 58
5.11 Definition of a viewpoint by distance and two angles – tilt and azimuth. 61
5.12 Occlusion level as detected in all images of the database. 63
5.13 Images and detected obstacles. Top - high occlusion level. Bottom - low occlusion level. 64
5.14 Relative total cost decrease. 65
5.15 Ratio of cases in which a second viewpoint was employed. 66
6.1 Illustration of target-sorting methods for sensing point y_1 and the revealed targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$. Left: optimal target sorting method. Right: Heuristic target sorting method (top-down). 73
6.2 Illustration of sensing methods for two sensing points y_1 and y_2 and the revealed targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$, $x_{RF}(y_2) = \{x_6, x_7, x_8\}$. A top-down target-sorting is used. Top-left: a-priori sensing. Top-right: batch sensing, Bottom: Sensing in harvest. 74
6.3 Robot acquisition of target locations and sensing points in the greenhouse as well as axis origin and orientation in the robotic base. 76
6.4 Detectability rate example marking Left: 50%; Right: 25%. 78
6.5 Simulated environment of the Kuka LWR4 robot and the targets. 79
6.6 Laboratory reconstruction of the greenhouse conditions on a KUKA LBR iiwa. . 79
6.7 Average travel cost per fruit as a function of sensing sequencing methodology and location probability function. 80
6.8 Average travel cost per fruit as a function of revelation probability and location probability function. 80
6.9 Average travel cost per fruit as a function of revelation probability and sensing methodology. 82
6.10 Average travel cost per fruit as s function of sensing methodology and joint detectability. 82
6.11 Total cost as a function of sequencing and sensing methodology. 83
7.1 Overview waypoint W_0 (camera facing front direction) and approach waypoint W_1, identified using the distance to fruit d and approach direction θ'_1. 88
7.2 Flowchart describing the two different approach strategies. Left: single approach direction; Right: multiple approach direction (differences marked with dashed lines). 90
7.3 For each target fruit i and potential approach direction θ'_j the control unit calculates the path of the robotic manipulator to a waypoint $W_{ij}(x'_{ij}, y'_{ij}, z'_{ij})$ in the multiple approach strategy. The figure illustrates a center-right-left approach. 91

7.4 The experimental setup in the greenhouse (left) compared to the previous experiment in laboratorial conditions using artificial fruits and leafs (right). . . 92
7.5 An overview image taken from the robot's camera looking at a laboratorial scene (left) and a greenhouse scene (right). 94
7.6 The position of each pepper was measured by manually moving the robotic arm close to the fruit and using the laser to estimate the distance to the fruit (the end-effector used in the greenhouse lacked the suction cup seen in the picture). 94
7.7 Example lab image of a fruit from three different viewpoints given to the questionnaire participants. They were told that the left and right viewpoints are reachable, while the center one is not. 96
7.8 An overview image from a greenhouse scene. The questionnaire participants were asked to determine all possible approach directions for the peppers marked with red boxes. 97
7.9 Rate of successfully approach attempts by approach direction θ_i', and the number of times each approach direction was attempted. In the right bars, cases where the robot failed to plan or collided with other fruits in the scene were removed. 98
7.10 Average number of approaches till successful fruit approach for the different potential approach directions of the multiple approach strategy described in Table 7.1 for two visual servoing distances. 100
7.11 Success rate for the different approach strategies described in Table 7.1 for two different approach distances. 100
7.12 Ratio of all approach attempts outcomes for the two approach strategies separately. Lost in VS means the fruit was no longer detected while doing visual servoing. Failure occurred due to not finding a safe path or due to collision between the end-effector and a fruit. 101
7.13 Interaction plot of fruit approach cycle time as function of approach strategy for reachable and unreachable fruits. 102
8.1 Robot location within the lane. d is the distance of advancing along the lane. . 109
8.2 High occlusion level crop. 109

List of Tables

1.1 List of gaps that are addressed in this thesis. 10
3.1 Summary of methods and corresponding chapters. 18
4.1 Databases description 39
4.2 Viewpoint description for each database. 39
4.3 Detectability results, single viewpoint by distance. 42
4.4 Detectability results, single viewpoint by tilt angle. 42
4.5 Detectability results, single viewpoint by azimuth angle. 42
4.6 Detectability results, single viewpoint by azimuth angle and lane side for DB#2, DB#3. For DB#2 lane sides (1) West (2) East. For DB#3 lane side (1)Southeast (2) Northwest. 43
4.7 Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and distance for DB#3. 43
4.8 Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and tilt angle for DB#3. 43
4.9 Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and azimuth angle for DB#3. . . . 44
4.10 Detectability results, combined viewpoints for DB#1. 45
4.11 Detectability results, combined viewpoints for DB#1 by distance, tilt and azimuth. 45
4.12 Modeling results, including significant variables and interactions, and regression weights. 46
4.13 Prediction model accuracy and significant variables at random train-test split. . 46
4.14 Prediction model accuracy and significant variables at temporal train-test split, DB#1. Average values across all possible combinations. 47
6.1 DB#1 – Generated scenes according to the axis in Figure 6.3. 77
6.2 DB#2 - 12 sensing points' description of locations and orientation. 77
6.3 Average decrease in travel cost for "sensing-in-harvest" sensing methodology compared to other sensing methods as function of target probability functions. 81
6.4 Average decrease in travel cost 60% revelation probability compared to lower revelation probability as function of target probability functions. 82
6.5 Average decrease in travel cost for "sensing-in-harvest" sensing methodology compared to other sensing methods as function of revelation probability. 82
7.1 Experimental protocol in the greenhouse compared to the laboratory experiments presented in [Ringdahl et al., 2017]. 93
7.2 Average fruit approach cycle times in seconds for the different approach strategies. 98
7.3 Average fruit approach cycle time in seconds for the different approach strategies described in Table 7.1 for two different visual servoing distances. 99
7.4 Fleiss' Kappa values for data collected in the laboratory and greenhouse. 102

Abstract

This thesis provides contributions to improve an autonomous harvesting robot's operational aspects, focusing on dynamic sensing and planning, leading to improved performance – increased detection and reduced cycle times. These gaps have been identified as major bottlenecks in the economic feasibility of current robotic harvesters. Four challenges have been addressed to optimize the system's performance – detectability and viewpoints analyses, dynamic sensing, task sequencing and fruit approaching. The case study of autonomous sweet-pepper harvesting has been selected for demonstration. All challenges are addressed using data acquired through a methodological acquisition of greenhouse data developed as part of this thesis. Using this methodology large datasets have been gathered in cooperation with other researchers as part of the EU Horizon 2020 SWEEPER project in which the results of this research has been partially implemented.

Detectability and multiple viewpoints modelling. Statistical models for fruit detectability were developed to provide insights into preferable variable configurations for better robotic harvesting performance. The modelling enables to identify how many peppers can be detected from a single sensing operation (viewpoint) or a combination of viewpoints. The methodology includes several steps: definition of controllable and measurable variables, data acquisition protocol design, data processing, definition of performance measures, and statistical modelling procedures. Given the controllable and measurable variables, a data acquisition protocol is defined to allow adequate variation in the variables, and determine the dataset size to ensure significant statistical analyses. Performance measures are defined for each combination of controllable and measurable variables identified in the protocol. Descriptive statistics of the measures allow insights into preferable configurations of controllable variables given the measurable variables' values. The statistical model is performed by back-elimination Poisson regression with a loglink function process. Spatial and temporal analyses are performed. The methodology was applied to develop statistical models for sweet pepper (*Capsicum annuum*) detectability revealing best viewpoints. 1312 images acquired from 10 to 14 viewpoints for 56 scenes were collected in commercial greenhouses, using an eye-in-hand configuration of a 6-DOF manipulator equipped with an RGB sensor and an illumination rig. Three databases from different sweet-pepper varieties were collected along different growing seasons. The conclusions were that target detectability greatly depends on the imaging acquisition distance and the sensing system tilt. A minimum of 12 training scenes are necessary to discover the statistically significant spatial variables. Better prediction was achieved at the beginning of the season with slightly better prediction achieved in a temporal split of training and testing sets.

Adaptive and dynamic sensing. Since each sensing operation costs time, the number of sensing operations must be minimized to ensure reduced cycle times. Dynamic sensing strategies were developed to improve detection results for a pepper harvesting robot. The algorithm decides if an additional viewpoint is needed and selects the best-fit viewpoint location from a pre-defined set of locations based on the predicted profitability of such an action. The suggestion of a possible additional viewpoint is based on image analysis for fruit and occlusion level detection, prediction of the expected number of additional targets sensed from that viewpoint, and final decision if choosing the additional viewpoint is beneficial. The

developed heuristic was applied on 96 greenhouse images of 30 sweet peppers and resulted in up to 19% improved detection. The harvesting utility cost function decreased by up to 10% compared to the conventional single viewpoint strategy.

Task sequencing. Given the decision on an additional viewpoint and its location, received from the dynamic sensing algorithm, the planner decides on how to sequence the sensing operation with the harvesting tasks in a time efficient manner. Planning the sequence of harvesting and sensing tasks is achieved using the traveling salesman paradigm by considering the costs of the sensing and harvesting actions along with the traveling times. The developed methodology is validated and evaluated in both laboratory and greenhouse conditions for a case study of a pepper harvesting robot. The results indicate that planning the sequence of tasks for a sweet pepper harvesting robot can reduce travel costs on average by 12% compared to currently used ordering heruristics.

Fruit approaching. After detecting the fruit, it must be approached for harvesting. Robotic harvesters that use visual servoing must choose the best direction from which to approach the fruit to minimize occlusion and avoid obstacles that might interfere with the detection along the approach. This work proposes different approach strategies, compares them in terms of cycle times, and presents a failure analysis methodology of the different approach strategies. The different approach strategies are: in-field assessment by human observers, evaluation based on an overview image using advanced algorithms or remote human observers, or attempting multiple approach directions until the fruit is successfully reached. In the latter approach, each attempt costs time. Alternatively, a single approach strategy that only attempts one direction can be applied if the best approach direction is known a priori. The different approach strategies were evaluated for a case study of sweet pepper harvesting, in laboratory and greenhouse conditions. The first experiment, conducted in a commercial greenhouse, revealed that the fruit approach cycle time increased 8% and 116% for reachable and unreachable fruits, respectively, when the multiple approach strategy was applied, compared to the single approach strategy. The second experiment measured human observers' ability to provide insights to approach directions based on overview images taken in both greenhouse and laboratory conditions. Results revealed that human observers are accurate in detecting unapproachable directions while they tend to miss approachable directions.

Automatic data acquisition protocols. Automatic data acquisition protocols were developed to supporting development and evaluation of the four main challenges. The data collected in the thesis served as a basis for algorithm development and validation, and has been partially released to the public to use as part of the H2020 SWEEPER project data management plan.

Keywords: autonomous harvesting; agricultural robots; robotic harvesters; dynamic sensing; adaptive sensing; approach planning; task planning; sweet pepper,agricultural databases; viewpoint;

List of Publications

The following publications resulted as part of this thesis contributions:

Journal publications:

- **Polina, Kurtser** & Yael Edan (2018). “Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers”. Biosystems Engineering, pp. 272–289.
- Ola Ringdahl, **Polina Kurtser**, & Yael Edan (2018). “Evaluation of approach strategies for harvesting robots: Case study of sweet pepper harvesting”. Journal of Intelligent and Robotic Systems, Springer, pp. 1-11.

Peer-reviewed conference proceedings:

- **Polina Kurtser** & Yael Edan (2018). “The use of dynamic sensing strategies to improve detection for a pepper harvesting robot”. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
- Ola Ringdahl, **Polina Kurtser**, & Yael Edan (2017). “Strategies for selecting best approach direction for a sweet-pepper harvesting robot”. Towards Autonomous Robotic Systems: 18th Annual Conference. Guildford, UK: Springer, pp. 516–525.

Conference presentations (without proceedings):

- **Polina Kurtser**, Boaz Arad, Ohad Ben Shahar, Milan van Bree, Joep Moonen, Bart van Tujil, & Yael Edan (2016). “Robotic data acquisition of sweet pepper images for research and development”. The 5th Israeli Conference on Robotics, 13-14 April 2016. Hertzilya, Israel.

In preparation and submission:

- **Polina Kurtser**, Yael Edan. "Planning the sequence of tasks for harvesting robots". Submitted to a robotics journal in 2018.

Additional publications related to this thesis (dynamic sensing and planning) in which I had major contributions:

- Elie Zemmour, **Polina Kurtser**, & Yael Edan (2017). “Dynamic thresholding algorithm for robotic apple detection”. Autonomous Robot Systems and Competitions (ICARSC), 2017 IEEE International Conference on. IEEE, pp. 240–246
- Boaz Arad, **Polina Kurtser**, Ehud Barnea, Ben Harel, Yael Edan, & Ohad Ben-Shahar. "Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper harvesting robots". Submitted.
- Elie Zemmour, **Polina Kurtser**, & Yael Edan. "Automatic parameter tuning for adaptive thresholding in robotic fruit detection". In preparation.

Related technical reports submitted as part of the Horizon 2020 SWEEPER project:

- **Polina Kurtser**, Boaz Arad, Yael Edan, & Ohad Ben-Shahar (Aug 2018). "D5.12: Final report on fruit detection, localization and maturity algorithms".SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- **Polina Kurtser**, Boaz Arad, Yael Edan, & Ohad Ben-Shahar (Aug 2018). "D5.11: Advanced software for fruit detection, localization and maturity, linked with results from WP6: optimized crop #2". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- **Polina Kurtser**, Boaz Arad, Ben Harel, Yael Edan, & Ohad Ben-Shahar (Oct 2017). "D5.9: Advanced software for fruit detection, localization and maturity, linked with results from WP6: optimized crop #1". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- **Polina Kurtser**, Boaz Arad, Ben Harel, Yael Edan, & Ohad Ben-Shahar (Sep 2017). "D5.10: Image databases #4 with different crop conditions available for analyses". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- **Polina Kurtser**, Boaz Arad, Ben Harel, Yael Edan, & Ohad Ben-Shahar (Jan 2017). "D5.6: Advanced software for fruit detection, localization and maturity". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- Ohad Ben-Shachar, Yael Edan, **Polina Kurtser**, &Boaz Arad (Sep 2016). "Report on evaluation methodologies". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- Boaz Arad, Efrat Taig, **Polina Kurtser**,Yael Edan, & Ohad Ben-Shahar (Sep 2016). "D5.4: Report on basic fruit detection, localization and maturity algorithms". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- Boaz Arad,Efrat Taig, **Polina Kurtser**, Ola Ringdahl, Peter Hohnloser,Thomas Hellström, Yael Edan, & Ohad Ben-Shachar (Sep 2016). "D5.2: Basic software for fruit detection, localization and maturity". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.
- Boaz Arad,**Polina Kurtser**,Yael Edan, & Ohad Ben-Shahar (Aug 2015). "D5.10: Image databases #1". SWEEPER, Sweet Pepper Harvesting Robot, Grant Agreement Number 644313.

Introduction

1

> *There can be economy only where there is efficiency.*
>
> **— Benjamin Disraeli**
> (Former British Prime Minister)

Current state of the art in autonomous harvesting robotics has demonstrated proof of concept in field conditions with operational feasibility achieved [Pitla, 2018; Blackmore, 2016; Bac et al., 2014b]. The goal set by researchers in the early 90's of solving the whole loop needed for successful robotic harvesting – crop detection, maturity evaluation, reaching the fruit, grasping and detaching it [Sarig, 1993] - has been accomplished for many case studies. Several of those robots (e.g., [McCool et al., 2016; Sa et al., 2016; Bac et al., 2014b; Hemming et al., 2014a; Kamilaris & Prenafeta-Boldú, 2018; Bac et al., 2017; Han et al., 2012; van Henten et al., 2002]) have presented very advanced capabilities, bringing the day where autonomous harvesting robots will be available on the market closer than ever. Several recent publications predict vast penetration of robotics into the agricultural field [Duckett et al., 2018; Pitla, 2018; Shamshiri et al., 2018]. These predictions impose that it is no longer enough to address the feasibility of the technologies, but these robots should be improved to perform quickly and efficiently. New building blocks must be approached to bring commercial profit to the growers and thereby lead to market penetration [Blackmore, 2016].

This thesis focuses on some of these blocks aiming to improve the performance of current autonomous harvesters. The main gap addressed in this thesis is improving the robot's operational aspects, focusing on adaptive sensing and planning leading to improved performance – increased detection and reduced cycle times. These have been identified as major bottlenecks in the economic feasibility of current harvesters [Bac et al., 2014b].

Current robotic harvesters work hierarchically in a sense-plan-act paradigm [Kim et al., 2005; Murphy, 2000]. The dynamically changing agricultural environment is unstructured [Bac et al., 2014b] and with three types of variation: variation of the object within the crop, the environmental variations, and variation between different types of crops [Bac et al., 2014b]. One way to overcome these problems is by shifting the paradigm into an operational flow where sensing and planning are interconnected and continuously repeated until an action is performed, by incorporating dynamic and adaptive sensing and task planning. This thesis proposes adaptive sensing and planning strategies and algorithms to overcome the uncertainties that are formed by the unstructured and highly variable conditions. Specifically this thesis focuses on the following four challenges: detectability and viewpoints analyses, adaptive/dynamic sensing, task sequencing and fruit approaching. Some of the solutions include development of new algorithms and techniques, while others deal with implementation and adaptation of methods used in other domains to agricultural robotics.

Since most agricultural domains are unstructured, robotic performance must be evaluated for a variety of crop conditions [Edan & Miles, 1994]. Therefore, to evaluate the performance of the algorithms as well as for their development, large datasets were acquired. To achieve this, autonomous robotic acquisition protocols were developed.

To demonstrate the applicability in real world conditions, the case study of autonomous sweet-pepper harvesting has been selected. The results have been partially implemented as part of the EU HORIZON 2020 SWEEPER project[1], aiming to bring a sweet pepper harvesting robot to the market. The results were published in four reviewed publications (two journal publications and two conference proceedings) and presented in three international conferences as detailed in the publication list (p. xxi). The developed methods, with some adjustments, can be applied to other harvesting robots, as described in the conclusions (Chapter 8).

1.1 Problem formulation

A robot, in general, has three functional components [Kim et al., 2005; Murphy, 2000]: sensing, planning, and acting. Several robotic paradigms were developed during the years to describe the relation between these three components [Murphy, 2000]. The hierarchical (deliberative) paradigm assumes that the robot operates top-down in a sequential mode, relying heavily on planning: the robot senses the world, plans the next action, and then acts. The reactive (behavior-based) and hybrid paradigms have been developed to tie together these functional components [Murphy, 2000]. The reactive paradigm assumes that actions are linked directly to sensing without planning them, while the hybrid paradigm assumes that the global tasks are planned and divided into sub-tasks that may be solved in a reactive manner.

As noted, the current practice for agricultural robots to date is the sense-plan-act paradigm. This thesis aims to tie together the sensing and acting functional components via dynamic sensing and task planning, a practice that has been developed for robots dealing with dynamic environments. While not proposing a new paradigm, the proposed algorithms focus on the ability to plan the sensing actions in a dynamic manner for agricultural harvesting robots. Planning of the sensing tasks is achieved by evaluating the costs of sensing in relation to their contribution to the overall performance. This planning contributes to improved detection while taking into account cycle times to ensure improved performance.

For simplicity, let us consider the following example. A robotic harvester has performed a sensing operation in which n targets were detected. The sensing action is completed and this information has been transferred to the planner. Current existing robotic harvesters that operate in a hierarchal/sequential manner would prioritize one of the fruits based on some sequencing mechanism and plan an approach and harvest action for that fruit and then move back to the sensing location to search again for the remaining $n-1$ fruits. Some harvesters will attempt to harvest the remaining $n-1$ fruits without sensing again by planning a harvesting sequence for all n fruits together. Now, let's assume a more advanced planner, which is based on previous knowledge of similar scenes. Based on this knowledge it can assume that m additional fruits have not yet been detected. Based on the hierarchical paradigm the planner will attempt to find them in the next sensing operations only after all first n fruits are harvested. The proposed dynamic paradigm (Figure 1.1) executes the additional sensing operation to be performed before an action is taking place, i.e., it plans the sensing

[1] http://www.sweeper-robot.eu

while performing the action. This planning is dependent on the scenario, hence requiring dynamic response from the robot. In the described example a planner may decide to perform an additional sensing immediately or after harvesting some of the targets, depending on their locations and on the actual number of fruits detected in each sensing operation.
The proposed operation flow (Figure 1.1) imposes the following new challenges which are the main contributions of this thesis (Chapters 4-7):

- **Detectability and multiple viewpoints modeling**. To be able to identify how many peppers can be detected from a single sensing operation (viewpoint) or a combination of viewpoints, modeling of detectability is required. Most harvesters are equipped with a RGB sensor placed on the robotic manipulator in an eye-in-hand configuration for fruit detection. Current best reported results are 87% detection rate with a false alarm rate of 3.8% [Kamilaris & Prenafeta-Boldú, 2018; Vitzrabin & Edan, 2016a; Bac et al., 2014b]. These performance measures are calculated using annotation where the labels are used as a benchmark for the algorithm evaluation. However, the annotation of a single image is not a sufficient benchmark for ground truth evaluation – the actual number of objects within the scene. For example, a scene containing 2 fruits, one fully visible on the image and one occluded by a leaf will be annotated by only segmenting the visible pepper. A detection algorithm that will be able to detect the labeled pepper but not the occluded one will yield a detection rate of 100%, but in reality only 50% of the peppers have been detected. Previous research [Hemming et al., 2014b; Bulanon et al., 2009] revealed that only about half of the fruit present in the field of view are visible from a single viewpoint. Hence, multiple viewpoints are necessary to overcome the detection rate bottleneck limiting harvesting performance. This thesis addresses a more profound analysis of fruits visible from each viewpoint, and the characteristics of the preferable viewpoints needed (Chapter 4), resulting in a new methodology for visibility analyses. The methodology is based on intensive data acquisition used for development and validation, as apposed to previous research which was done on limited manually acquired datasets.

- **Dynamic and adaptive sensing**. Given the understanding that often more than a single viewpoint is needed to increase overall detectability, one would suggest an infinite number of viewpoints to reach maximum detectability. This would have been possible in a non-resource constraint environment where the time given to the robot to detect a fruit is infinite. In a more realistic scenario, where the harvesting cycle time is a major bottleneck preventing commercialization of harvesting robots [Bac et al., 2017; Elkoby, 2016; Elkoby et al., 2014], the number of viewpoints should be limited, and the location of the next viewpoint should be carefully calculated so as to select viewpoints that provide maximum detectability is applied.
  With the general guidelines of preferable viewpoints or combinations of viewpoints, gathered in the detectability analysis, the planner must decide if an additional viewpoint is needed based on the information retrieved from the first sensing operation in the first viewpoint. Then, the cost of an additional viewpoint must be weighed against the

predicted profit from sensing additional targets and the location from which to make another sensing operation must be derived. In this challenge active sensing methods are introduced to the agricultural robotics domain while developing novel information content measures that are most suitable to the proposed environment (Chapter 5).

- **Task sequencing**. Given the decision on an additional viewpoint and its location, received from the adaptive sensing algorithm, the planner must decide how to sequence the sensing operation with the harvesting tasks in a time efficient manner. Since the robotic arm is equipped both with the sensors and the harvesting tool, both the sensing and the harvesting operations require traveling of the robotic arm between locations of sensing viewpoints and sensed targets. Reduction of the travel time required can significantly reduce the harvesting cycle time. Therefore, the third challenge of this thesis addresses the sequencing optimization problem. It aims to minimize the overall cycle time by optimizing the sequence of sensing operations and targets to be visited based on their spatial locations (Chapter 6).

- **Fruit approaching**. The first viewpoint from which a detection is performed is typically done from an overview location from where several plants are visible. Once a target is detected, approaching the fruit when eye-in-hand cameras are used is often divided into two steps [Barth et al., 2016]. The first step involves moving from the overview location to an approach location where a single fruit is centered in the image. The second step uses visual servoing [Barth et al., 2016; Chang, 2007] to move towards the fruit until it is reached and is necessary to continuously refine the fruit position detected from the overview image. Given the complexity of the environment and the continuous detection required when using visual servoing, there is a high risk of losing the fruit along the approach. Therefore, it is important to approach the fruit from an approach location where the fruit is not occluded by leaves and other obstacles along the approach. The fourth challenge addressed in this thesis (Chapter 7) deals with evaluation of different strategies for fruit approaching.

The main common thread tying together these challenges is the need **to plan the sensing actions to reach optimal detectability in minimum time (minimal arm movements)**. With each sensing action, detectability increases; however, with extra arm movement harvesting cycle time also increases and the profitability of the robot decreases. On the other hand, with each undetected fruit the efficiency of the robot decreases and so does its profitability.
To enable the development of all four challenges real-world data had to be acquired. Beyond the importance of validating algorithms and evaluating performance in real field conditions and on real world [Edan et al., 2000] the agricultural domain places extra benchmarking challenges due to the high variability and limited repeatability available [Edan & Miles, 1994]. Therefore, performance must be evaluated on a wide range of conditions and based on a statistical analysis [Edan & Miles, 1994] requiring large amounts of data. Protocols for automated data acquisition in greenhouse conditions are described in Chapter 3.2 and are a major contribution essential for all the algorithms development.

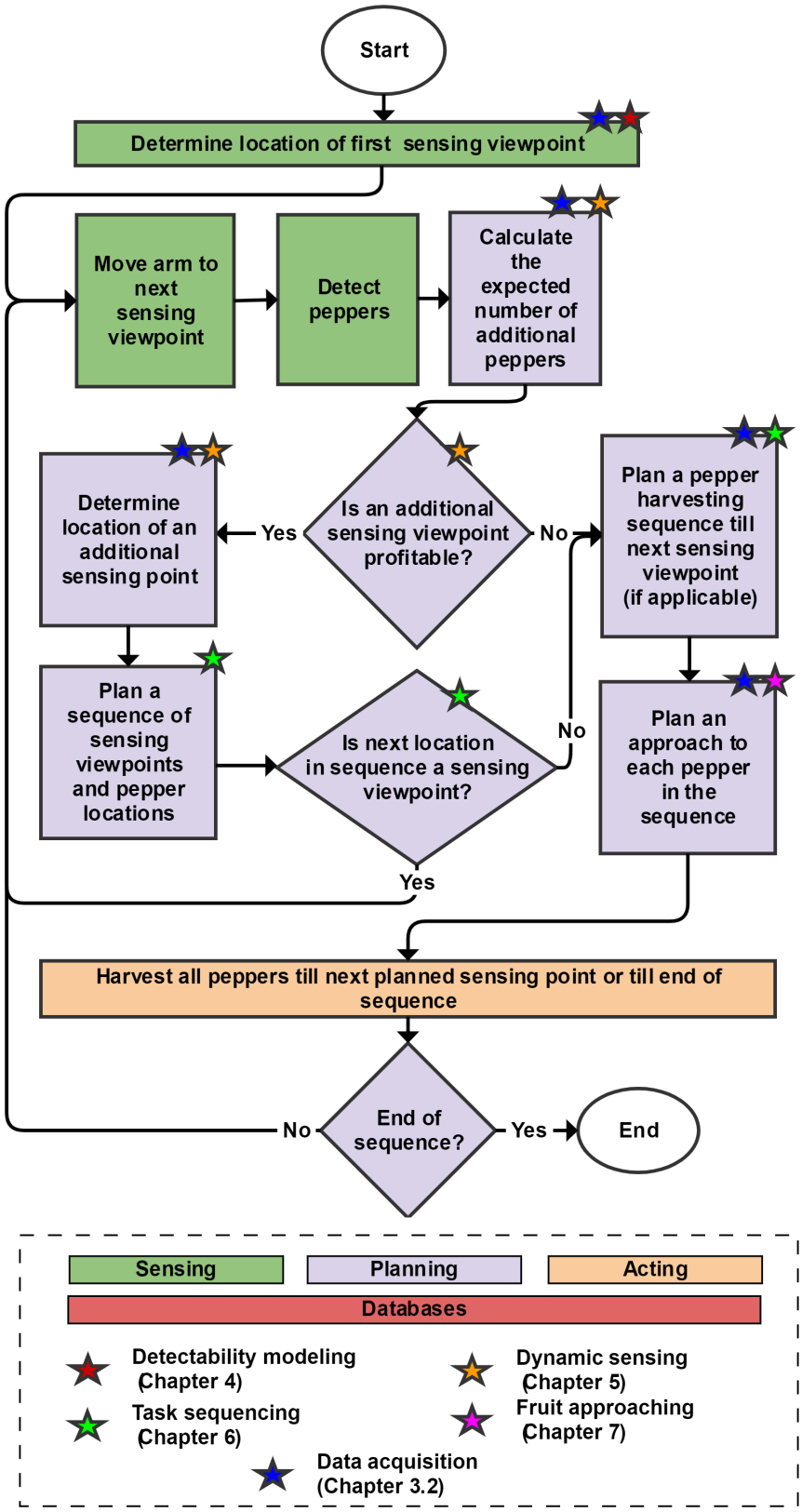


Figure 1.1: Autonomous pepper harvester sensing-planning-acting paradigm breakdown and the defined challenges.

1.2 Research objectives

This thesis outlines the following research objectives corresponding to the above noted challenges:

- **RO1: Statistical evaluation of detectability from a combination of viewpoints to provide insights into characteristics of best viewpoints.**
  - RQ1.1: What are the static best characteristics of a sensing viewpoint for sweet pepper harvesting? Are there any at all?
  - RQ1.2: How do the preferred viewpoint characteristics change along the season, growing conditions, cultivation techniques, and varieties?
  - RQ1.3: What methodologies should be applied to perform a similar analysis for different crop variety/conditions?
  - RQ1.4: What is the minimal dataset needed for dynamic learning of best viewpoint characteristics for a given setting?
  - RQ1.5: Are there temporal relations between subsequent plants along the row in terms of detectability and best viewpoint characteristics?
- **RO2: Development of an dynamic sensing algorithm that will predict the need for an additional viewpoint and its location.**
  - RQ2.1: How should information content of a viewpoint be measured?
  - RQ2.2: How to predict the overall number of peppers in the scene based on information gathered from one viewpoint?
  - RQ2.3: How to make a profitable decision on another viewpoint?
  - RQ2.4: Where should an additional viewpoint be placed? How is the location dependent on the information extracted from the first viewpoint?
- **RO3: Compare different strategies of harvesting and sensing sequencing.**
  - RQ3.1: How to calculate traveling distance/cost function given a robot configuration?
  - RQ3.2: What is the travel time decrease if optimization is introduced in the described conditions with full a priori knowledge of target locations?
  - RQ3.3: How should unknown locations be treated in planning of a harvesting sequence?
- **RO4: Compare different approach strategies.**
  - RQ4.1: What is the value of knowing the approach direction vs searching for it?
  - RQ4.2: Can humans identify an approachable direction?
  - RQ4.3: What common failures occur during fruit approach in visual servoing?

To evaluate all above research objectives in real world scenarios the case study of autonomous sweet-pepper harvesting was selected for demonstration. All challenges were addressed using data acquired through the following research objective which provided the basis for evaluation of all research objectives:

- **RO5: Data collection protocols in agricultural settings and collection of large agricultural databases.**

1.3 Innovation and contribution

Autonomous robotic harvesters are now on a verge between proof-of-concept products to machinery that can actually enter the market. Addressing optimization problems that shorten the harvesting cycle time and increase detectability rate is a “hot topic” that is in its first steps of development [Blackmore, 2016]. The idea of planning the sensing actions in order to achieve these two goals is the main innovation and contribution of this thesis. The thesis provides new solutions to several challenges that have not been approached or have been approached in very limited contexts in the scientific literature of agricultural robotics – all related to adaptive sensing and planning. The specific gaps this thesis addresses are noted in Table 1.1.

Specifically, this thesis contributes with the following innovations. First, for developing and evaluating algorithms real world data is needed. This is especially important in the unstructured agricultural environment due to the high variability and limited repeatability available [Edan & Miles, 1994]. Therefore, performance must be evaluated on a wide range of conditions and based on a statistical analysis requiring large amounts of data. A common practice for R&D of complicated problems is the acquisition of a large database (e.g., Labelme [Russell et al., 2008]). These datasets enable to advance algorithms development [Sa et al., 2016; Kapach et al., 2012] and provide a benchmark for evaluating new algorithms. To the best of our knowledge, to date there is no large dataset available for R&D dealing with agricultural objects in the greenhouse environment. A major reason for the limited data is the manual acquisition protocols applied that require tedious high labor demands. In this thesis we developed and implemented an automatic data acquisition protocol (Section 3.2). The thesis is based on field experiments that include a robotic arm operating in the greenhouse; as result of the automatic data acquisition a considerable number of labeled images were collected and made available to the public as part of the data management plan of the SWEEPER project[2]. The collected data has been used in each of the algorithms developed as described in Chapters 4-7 and serve an important basis for evaluating the algorithms developed on real world data.

Second, the question of detectability from a single or multiple viewpoints has been limitedly covered before [Hemming et al., 2014b; Bulanon et al., 2009]. Previous research analyzed single viewpoint detection rate (number of fruits visible from a viewpoint with ideal detection algorithm due to occlusions) for peppers [Hemming et al., 2014b] and citrus plants [Bulanon et al., 2009]. Their conclusions were based on a limited amount of data due to manual acquisition, and in un-generalized conditions (all were collected in the same greenhouse, same variety, single repetition in a specific growing season). This thesis (Chapter 4) introduces a statistical methodology for analysis of viewpoint data on a larger dataset that includes variability in season, growing conditions, and varieties. It also includes analysis of sequential

[2]The specific acquisition protocol has been designed specifically for the data collection required for this thesis and was conducted by me for the datasets analyzed in this thesis; the actual databases are a combined effort of several researchers in this project - it required special programming of the robotic and imaging systems; all databases noted in this thesis have been collected by me (with technical support of others).

temporal relations that has not been addressed in previous studies. The contribution of the analysis is not only in drawing conclusions for the conditions described, but also in providing a methodology for future similar analysis by automating the procedure, which allows a larger dataset to be analyzed in a relatively fast manner. The results have been published in the Journal of Biosystems Engineering [Kurtser & Edan, 2018a].

Third, integration of dynamic sensing into the domain of agricultural robotics has not been previously considered in any of the literature reviewed. Dynamic and adaptive sensing methodologies have been explored in other domains in highly predictable environments where the sensing is based on heavy modeling of the environment (see more details in background section in Chapter 5.1). The lack of such a model, or the ability to develop one, as was shown in the statistical analysis in Chapter 4 [Kurtser & Edan, 2018a], limits the ability to use the standard measures of information content often used in dynamic sensing [Vázquez et al., 2002; Vázquez et al., 2001; MacKay, 1992; Bajcsy, 1988]. Furthermore, the vast majority of general dynamic sensing literature, providing methodologies for adaptive and dynamic sensing, overlooks the question of resource allocation and the cost of additional sensing. Since each sensing operation takes time (mostly due to arm movements to the new sensing location) and therefore prolongs the overall harvesting cycle time (Section 2.3), the challenge of dynamic sensing should also consider these sensing costs. The algorithms developed in Chapter 5 address both described challenges – the measurement of information content in the dynamic agricultural setting and the implementation of sensing costs for the dynamic decision support of an additional sensing operation. These are both new for agricultural harvesting robots. These results have been published in the Proceeding of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018) [Kurtser & Edan, 2018b].

Fourth, the application of operations research (OR) methodologies in the agricultural domain has been traditionally applied for improving logistics of greenhouse management (e.g., allocation of workers, heavy machinery, cycles of operations; see Section 2.4), but has not yet been considered in the autonomous harvesting process. The first contribution in this domain has been in the application of harvesting sequencing optimization using methodologies developed for solving the traveling salesman problem (Section 2.4). This includes development of suitable utility functions that describe the cost of travel between two points by a robotic harvester and evaluation of their performance. The development of a task planning algorithm that deals with sequencing of the sensing and harvesting tasks is addressed in Chapter 6. This research is underway to be submitted to a robotics journal.

A fifth contribution is related to the question of motion planning in robotic arm operation, which has been vastly explored in the general robotics literature with new algorithms appearing on a regular basis. These algorithms assume perfect knowledge of the location of the target to be reached and the obstacles in the scene are either known a priori or constantly sensed. In the cases where the targets' locations are not accurately known, a visual servoing mechanism [Barth et al., 2016] is applied where the fruit is being re-detected over and over again, adjusting its location until reached. Since the redetection is not necessarily performed from the location where the fruit has been detected in the first instance, there is a chance

of losing the fruit if it has been approached from an occluded direction. This thesis focuses on determining the correct approach direction. Strategies of approach direction determination are described in Chapter 7, including repetitive search and inclusion of a human in the loop. The results have been published in both a conference proceeding and a journal publication[3] [Ringdahl et al., 2018; Ringdahl et al., 2017].
All the above developments have been applied to real field data from the HORIZON 2020 SWEEPER project [*Horizon 2020 SWEEPER* 2015-2018] with some of the conclusions also partially implemented in the operational system. The data was collected with protocols defined as part of this thesis (described in Section 3.2); these protocols can be applied for collecting data for additional crops and evaluating the developed algorithms.

1.4 Thesis Structure

Chapter 2 outlines the background supporting the claims made in the introduction and a general literature review of the state of the art in the relevant fields for the proposed questions. Chapter 3 describes the common methods used to meet the challenges defined in the research objectives section (Section 1.2) including the data acquisition protocols and the availability of the gathered data. Chapters 4-7 describe each of the algorithms developed (detectability analysis; dynamic sensing; task planning; and fruit approaching). Each of Chapters 4-7 are published or submitted papers. Hence, each chapter is independent and includes background, methods, results, and conclusions sections. As a result there is some repetition in the introduction sections and some of the literature review in this thesis. Chapter 8 describes the overall conclusions drawn from this research, and suggestions how to adapt the methodologies to other crops and environments along with recommendations for future work.

[3] The idea to evaluate different approach strategies was my contribution, the specific strategies proposed and their implementation were a collaborative effort advanced in several iterations. Advanced development of software for the robotic manipulator was necessary to validate the proposed strategies. This was conducted mainly by the first author who was the project software integrator. I was involved in the software implementation throughout the process (design and implementation). The experimental design and data analysis were performed by me, while the field experiments and paper writing were a collaboration.

Table 1.1: List of gaps that are addressed in this thesis.

Problem	Introduction and background sections	Gap	Thesis contributions (Chapter)	
Acquisition of large databases for algorithm development	1.1.1 - Detectability and multiple viewpoint modeling 2.2 - Fruit detection from single and multiple viewpoints	Current algorithms in harvesting robotics are evaluated on small manually collected dataset	Acquisition using a robotic harvester. Development of protocols and making data available to the public	Chapter 3 (sections 3.2-3.3)
Best viewpoint characteristic for detection	1.1.1 - Detectability and multiple viewpoint modeling 2.2 - Fruit detection from single and multiple viewpoints	The characteristics of a sensing viewpoints has not been investigated on large datasets	Statistical modeling of spatial and temporal relations between scenes	Chapter 4
Next best viewpoint search	1.1.1 - Detectability and multiple viewpoint modeling 1.1.2 - Dynamic and adaptive sensing 2.2 - Fruit detection from single and multiple viewpoints 2.3 - Harvesting cycle times	The search of next best viewpoint has not been done in agricultural robotics	Dynamic sensing through statistical estimation of number of fruits and cost of another view	Chapter 5
How to order the tasks of sensing and harvesting	1.1.3 - Task sequencing 2.3 - Harvesting cycle times 2.4 - Traveling salesman problem in agriculture	No optimization of harvesting sequencing or sensing actions is done in agricultural robotics	Introduction of traveling salesman problem (TSP) with specially made utility function and addressing unknown targets	Chapter 6
How to approach a detected pepper without losing it from sight	1.1.4 - Fruit approaching	After detection the path planning to the fruit doesn't take into account the least occluded direction	Introduction of a search pattern or a human in the loop.	Chapter 7

Background

2

> *The chances of finding out what's really going on in the universe are so remote, the only thing to do is hang the sense of it and keep yourself occupied.*
>
> **— The Hitchhiker's Guide to the Galaxy**
> (Douglas Adams)

This chapter provides the background related to autonomous harvesters challenges and bottlenecks. The overview includes review of:

1. Robotic harvesting & fruit detection algorithms,
2. Viewpoint analysis,
3. Harvesting cycle times calculations and estimations,
4. Implementation of the traveling salesman problem into the agricultural domain, and
5. Available agricultural datasets.

More detailed reviews related to each challenge are described in the introduction section of each chapter related to each algorithm developed.

2.1 Autonomous robotic harvesters

Robotic harvesters operating autonomously in the field has been a vision for the precision agriculture community since the mid-eighties [Sarig, 1993; Edan et al., 1991; Edan et al., 1993]. The tedious manual labor involved in the harsh agricultural domain has led to shortage in manpower leading to rising demand for automation of the processes involved. This need has brought automation to almost all elements of greenhouse operation. In today's cutting-edge greenhouses all sorting, packaging, and fruit transportation is performed automatically with the human serving only as a supervisor. Harvesting is one of the few tasks that have not yet been automated and is still performed manually.

Research in harvesting robots [Bac et al., 2014b] focuses on robot design (e.g.,[Bloch et al., 2017; De-An et al., 2011; Edan & Miles, 1994; van Henten et al., 2009]), sensing [Kamilaris & Prenafeta-Boldú, 2018; Kapach et al., 2012; Zhao et al., 2016; Sa et al., 2016], path planning [Luo et al., 2018; Bac, 2015; Barth et al., 2016] and grasping (e.g.,[Eizicovits & Berman, 2014; Eizicovits et al., 2016]). Currently reported attempts of fully integrated autonomous harvesting includes cucumber harvesting [van Henten et al., 2002; van Henten et al., 2003], strawberry harvesting [Hayashi et al., 2010], cherry tomatoes [Tanigaki et al., 2008; Feng et al., 2018], eggplant harvesting [Hayashi et al., 2002], apple picking [De-An et al., 2011; Yuan et al., 2016], and sweet peppers harvesting [Bac et al., 2017; Lehnert et al., 2017].

Robotic harvesters are often equipped with an eye-in-hand configuration, where an RGB camera is mounted at the tip of a robotic manipulator[Bac et al., 2014b]. Some robotic harvesters are also equipped with additional sensors, such as an RGB-D camera or a second camera (used for stereo vision) to be used on both the robotic manipulator and the robotic cart placed statically within the greenhouse lane [Bac et al., 2014b; van Henten et al.,

2002; Barth et al., 2016; Bontsema et al., 2015]. To allow sensing operations in various lighting conditions, some robotic harvesters are equipped with artificial illumination, placed on the cart [Bontsema et al., 2015; Hemming et al., 2014a] or on the manipulator [*Horizon 2020 SWEEPER* 2015-2018]. Artificial illumination allows both acquisition in low intensity natural light condition as well as development of advanced algorithms for improvement of object detectability, (e.g., Flash-No-Flash [Arad et al., 2018]). These algorithms use both the illuminated and the natural light images to cope with intense natural light. Others apply advanced algorithms for dynamic thresholding (e.g., [Vitzrabin & Edan, 2016a; Zemmour et al., 2017; Ostovar et al., 2018]).

In their 2014 review, which served as a major motivator for this thesis, Bac et al. [Bac et al., 2014b] outlined the main bottlenecks preventing autonomous harvesting from penetrating into the market. The first bottleneck is the low detection rates of current vision algorithms. The reported numbers in the review were 87% detection rate and since 2014 this has not increased significantly [Kamilaris & Prenafeta-Boldú, 2018; Zhao et al., 2016]. The second bottleneck is the too long harvest cycle times, reducing the profitability of the robot. This thesis contributes in developments that aim to improve both bottlenecks by developing dynamic sensing and task planning algorithms to improve the harvesting process.

2.2 Fruit detection from single and multiple viewpoints

Many algorithms and methods have been developed for fruit detection as noted in several reviews [Kamilaris & Prenafeta-Boldú, 2018; Zhao et al., 2016; Kapach et al., 2012] and a few more relevant papers [Bargoti & Underwood, 2017; Zemmour et al., 2017; Sa et al., 2016; Vitzrabin & Edan, 2016a; Vitzrabin & Edan, 2016b; Gongal et al., 2015]. The most common method for fruit detection is the implementation of image segmentation algorithms used in other domains [Ostovar et al., 2018; Rong et al., 2017; Zemmour et al., 2017; Arroyo et al., 2016; Wang et al., 2013]. These algorithms include different methods such as K-means [Shmmala & Ashour, 2013]; mean shift analysis [Zheng et al., 2009]; artificial neural networks (ANN) [Al-Allaf, 2014]; support vector machines (SVM) [Sakthivel et al., 2015]; and deep learning [Kamilaris & Prenafeta-Boldú, 2018].

These methods, applied to agricultural datasets, report detectability rates up to 87% with a false alarm rate of 3.8%. To calculate the true detectability rate, the standard procedure applied in the computer vision community involves image labelling [Russell et al., 2008]. Manual annotators review an image [Deng et al., 2009; Russell et al., 2008] and segment it into areas that represent targets and background. These labels are often used both as a training set for supervised learning as well as a benchmark for detection algorithms.

However, the annotation of a single image is not sufficient for object detectability within a scene. The number of targets within the scene is defined as the ground truth. The common procedure for obtaining detectability ground truth is either manual counting of the number of fruits within the scene [Hemming et al., 2014b; Bulanon et al., 2009] or placing several sensors that simultaneously sense the same scene and then combining and annotating the

joint number of targets in the scene [Dollar et al., 2012; Russell et al., 2008]. Hemming et al. revealed that only 40-60% of the fruit present in the field of view are visible from a single viewpoint by a human observer [Hemming et al., 2014b]; this work was limited and based on analysis of 30 sweet pepper plants. Similar research on citrus harvesting [Bulanon et al., 2009] for a total number of 5 trees reported 40-70% visibility. Despite these low detection rates most work to date in agricultural robotics [Kamilaris & Prenafeta-Boldú, 2018; Bac et al., 2014b; Kapach et al., 2012; Bulanon et al., 2009] rely on detection from a single viewpoint. The main reason is probably that since the focus was on proving the technical feasibility of detection, R&D was concentrated on sensing methods and algorithms.
This question of dynamically calculating another viewpoint is addressed in the literature as the "best-next viewpoint" problem. The process of a best viewpoint search mostly depends on how much a priori geometrical information about the scene is available [Maver & Bajcsy, 1993]. The search for an optimal viewpoint has been extensively investigated in many fields, such as computational geometry [Vázquez et al., 2001; Maver & Bajcsy, 1993], graph drawing [Vázquez et al., 2001], and multi-camera stereo vision [Zabulis & Daniilidis, 2004]. In robotics applications optimal viewpoint analyses have been investigated for motion planning [Vázquez et al., 2001], grasping [Krainin et al., 2010], and medical intervention planning [Mühler et al., 2007]. These algorithms often rely on static placement of the sensors and a priori information. Since the greenhouse environment is very dynamic and unstructured, the best viewpoints cannot be selected a priori and therefore the selection of viewpoints should be adaptive to the sensed information. The selection of the viewpoint can not promise an optimal viewpoint, but a better one given the environment sensed. Dynamic selection of viewpoints is regarded as adaptive sensing, active sensing, active perception, dynamic sensing, or next best viewpoint selection algorithms. Classical research in the active sensing field [Bajcsy, 1988] assumes that: "in scenarios where data measurements are relatively expensive or slow, we want to know where to look next so as to learn as much as possible..." [MacKay, 1992], where information is often measured as viewpoint entropy [Vázquez et al., 2002].
Additionally, the question of visibility from a viewpoint has been addressed in the search of the least occluded direction to perform visual servoing from while planning a path towards the fruit. Related research dealt with determining reachability cones for robot design [Bloch, 2017] and obstacle detection (e.g., [Bac, 2015; Barth et al., 2016]).

2.3 Harvesting cycle times

In order for a robot to be commercially relevant, it should not only harvest peppers but also do so in a fast manner. An economic evaluation for sweet-pepper harvesting in the Netherlands [Pekkeriet, 2011] showed that a cycle time of 6s should be achieved. This number was provided for a robot catalog price of € 100k, 50% of the sweet peppers harvesting rate, use of 120h per week for 35 weeks per year within the harvesting season, and the assumed worker wage of € 16 per hour. In comparison a skilled human harvests peppers at 3.6 sec per fruit [Elkoby, 2016; Elkoby et al., 2014; Pekkeriet, 2011]. Current state of the art robots reported an average of 33 sec for all reviewed fruits [Bac et al., 2014b] with 94 sec per pepper in the latest paper by Bac et al. as part of the FP7 CROPS project [Bac et al., 2017]. Suggestions

to reduce cycle time [Bac et al., 2017] included: post-harvest logistics optimization (e.g., collection of fruits through a pipe versus carrying them to a collecting basket), better motion planning (e.g., waypoints used to reach the harvesting position), and reduction of number of motions required for sensing (e.g., use of a static sensor on the cart as opposed to an eye-in-hand configuration). While the use of a static sensor instead of an eye-in-hand configuration will reduce cycle times, due to the viewpoint analysis described in Chapter 4, it will limit the number of detected peppers to 40%-60% making the robot commercially unfeasible. Alternatively, the reduction of number of sensing points or unnecessary arm movements due to incorrect task and harvest sequencing can significantly reduce cycle times, a solution is proposed in this thesis (Chapters 5 and 6).

2.4 Traveling salesman problem in agriculture

The application of operations research (OR) in the agricultural domain was traditionally applied for improving logistics of greenhouse management (e.g., [Bochtis et al., 2015; Bochtis & Sørensen, 2010; Bochtis & Sørensen, 2009; Ali et al., 2009]), where an action, such as watering or seeding, is required to be accomplished by multiple vehicles/humans or cooperation between the two. Often the problems are distributed (e.g., coordination of teams of autonomous agricultural vehicles [Vougioukas, 2012]) and involve route planning [Edwards et al., 2017; Bochtis et al., 2015], or coverage planning [Jensen et al., 2015].

The optimization problems presented in this thesis are derived from the field of operations research with the specific relevant paradigm the traveling salesman problem (TSP). The TSP is a combinatorial optimization problem that has been widely researched and studied over the past few decades [Laporte, 1992; Lawler, 1985; Rosenkrantz et al., 1977], thereby giving rise to several formulations. The most widely used definition is the optimal tour solution of a "complete weighted undirected graph G, specified by a pair (N, d) where N is a set of nodes and d is a distance function mapping pairs of nodes (or edges) into real numbers" [Rosenkrantz et al., 1977]. The large-scale combinatorial problem is NP complete and is hence usually solved by a heuristic that provides a near optimal solution in reasonable computational times, depending on the application and computational power available.

The traveling salesman problem, or the vehicle routing problem, in the context of agricultural applications has been addressed in the literature for agricultural field logistics problems (e.g.,[Bochtis & Sørensen, 2010; Ali et al., 2009; Bochtis & Sørensen, 2009]), where an action, such as watering or seeding, is required to be accomplished by multiple vehicles or humans or cooperation between the two. The required action is repetitive and static, and an optimal plan should be developed before traveling, relying on ideal knowledge of the targets and the cost of transportation between them.

In fully integrated robotic harvesters to date, planning of the sensing operations has not been noted [McCool et al., 2016; Bac et al., 2014b; Hemming et al., 2014a]. The harvesting sequence is usually defined using a heuristic sequence with some pre-set order such as distance to the sensing rig or along the lane. Applying the solution of the traveling salesman problem (TSP) [Laporte, 1992; Lawler, 1985] to the case of harvesting robots has been noted to significantly decrease harvesting cycle times [Mann et al., 2016; Zion et al., 2014;

Edan et al., 1991]. Edan et al. showed that for a citrus harvesting robot the near-optimal harvesting sequence of fruits depends on the robot kinematics and the tree structure, by applying solutions of the TSP cycle times can be reduced [Edan et al., 1991]. Mann et al. [Mann et al., 2016] and Zion et al.[Zion et al., 2014] developed algorithms to define the sequence of melons picking for a multi-arm melon-picking robot by modeling the problem as a task of coloring an interval graph; they then used a greedy algorithm to produce an optimal solution for a k colorable sub-graph problem. In both these cases, the set of targets was assumed to be known in advance. The case when targets are unknown a priori and are discovered while traveling is called the rolling horizon TSP, [Golden et al., 2008] or the stochastic TSP also known as the "dynamic salesman problem". These methods introduce randomness and uncertainty into some of the initial conditions, or provide an adaptable to new information traveling plan [Pillac et al., 2013; Gendreau et al., 1996]. The distance utility function, aimed to be minimized by the TSP, should take under consideration the path planning capabilities of the robot as well as the mechanical configuration of the arm.
In this thesis we applied (Chapter 6) TSP to the dynamic case in which new target points are revealed along travelling to existing points corresponding to the defined research objectives.

2.5 Data acquisition

A common practice for image processing R&D for complicated problems is the acquisition of a large database (e.g., Labelme labeling database [Russell et al., 2008], Oxford building dataset [Philbin et al., 2007]). These datasets enable to advance vision algorithms development [Szeliski, 2010] and provide a benchmark for evaluating new algorithms.
To the best of our knowledge, to date there is no open dataset available for R&D in research related to agricultural objects. Evaluation of previously reported algorithms was based on limited data [Sa et al., 2016; Hemming et al., 2014b; Bulanon et al., 2009]. Previous research indicated the importance of evaluating algorithms for a wide range of sensory, crop, and environmental conditions [Hemming et al., 2014b]. Furthermore, real world data is important for algorithm development and validations. Moreover, since most agricultural domains are unstructured with high variability, robotic performance must be evaluated for a variety of crop conditions [Edan & Miles, 1994] and/or based on statistical analyses of vast amounts of data.

Research methods

3

> *Though this be madness, yet there is method in't.*
>
> — **Hamlet**
> (William Shakespeare.)

This chapter provides an overview of methods applied in the thesis. The methods used for developing each of the algorithms, are described in detail in Chapters 4-7.

3.1 Overview

The thesis contributes with development of methods related to the following gaps (Table 3.1, Figure 1.1):

1. Acquiring large databases.
2. Statistical analysis of viewpoint data.
3. Dynamic sensing.
4. Task planning.
5. Determining the approach direction.

The methodology for acquisring large databases in based on an **automatic data acquisition protocol**, described in Section 3.2. A robotic acquisition system and procedure was developed using a 6 degree of freedom manipulator, equipped with 3 different sensors to automatically acquire images from several viewpoints with different sensors and illumination conditions (see detailed description of hardware in Section 3.2). By using a robotic manipulator high precision and repetition are achieved allowing fast acquisition of greenhouse data along the day and the growing season. This allows greater homogeneity within the data that was acquired in similar conditions. On the other hand, it imposes challenges such as manual labeling and ground truth data acquisition. The pros and cons of this methods are also covered in this chapter.

Based on the data gathered according to the protocols defined in Chapter 3.2, in Chapter 4, I describe a methodology developed for **statistical analysis of viewpoint data**. This includes acquisition of three separate databases with high variability in both viewpoint features (angles, distances, focal length) and greenhouse features (varieties, growing methods, countries), allowing application of statistical methods for evaluating detectability from each viewpoint. Two viewpoints are considered different given any change in their features (change in angle, distance or focal length). By splitting the data into "training" and "testing" sets in both spatial and temporal fashions, the different relations between consecutive scenes are explored in a way of predictive models generation. The developed methodology allows insights into the preferable viewpoints and viewpoint combinations as well as identification of the number of viewpoints needed for a given crop. The chapter also outlines the differences between detectability and visibility and proposes alternative methods for ground-truth collection that are more suitable for mass acquisition, such as the robotic data acquisition proposed. The limitations of the method as well as the crops suitable for such an analysis are outlined.

Table 3.1: Summary of methods and corresponding chapters.

Problem	Gap	Methods	
Acquisition of large databases for algorithm development	Current algorithms in harvesting robotics are evaluated on small manually collected datasets.	Development of a protocol based on a 6DOF robotic manipulator equipped with color and depth sensors as well as artificial illumination. Acquisition of 5 datasets contributing to the thesis and additional 3 open access datasets, all in commercial greenhouses.	Chapters 3.2-3.3
Best viewpoint characteristic for detection	The characteristics of sensing viewpoints has not been investigated on large datasets	Acquisition of 3 datasets and analyzing them by finding statistically significant parameters through a Poisson regression model. Splitting the dataset in temporal manner allows analysis of spatial and temporal features. Testing on data gathered in greenhouse conditions.	Chapter 4
Next best viewpoint search	The search of next best viewpoint has been limited in agricultural robotics	Development of a dynamic next best viewpoint decision support algorithm for agricultural harvesting robot that analyzes a given image for prediction of number of targets from an additional viewpoints and compares the financial loss of those targets in comparison to the financial loss as a result of prolonged scanning action. Testing on data gathered in greenhouse conditions.	Chapter 5
How to order the tasks of sensing and harvesting	No optimization of harvesting sequencing or sensing actions is done in agricultural robotics	Employment of TSP for harvesting sequencing and evaluation of task planning strategies including those that sense first and then harvest all fruit or those optimizing the sensing location in the overall traveling sequence. Testing in laboratory and simulative conditions.	Chapter 6
How to approach a detected pepper without losing it from sight	After detection the path planning to the fruit doesn't take into account the least occluded direction	Approach strategies are evaluated for their efficiency (search pattern vs single approach vs human determining the approach direction).	Chapter 7

Given the conclusions of Chapter 4, that there are no single or combinations of viewpoints that can be predefined as the best ones in any operation mode, but only guidelines that depend on growing conditions (e.g., time in the season, location in the greenhouse), a **dynamic sensing algorithm** is developed in Chapter 5. This algorithm, decides if an additional viewpoint is needed and where should it be located based on the analysis of a single viewpoint image. It is achieved by estimating the expected number of peppers to be detected from an additional viewpoint using a Poisson regression model. Then comparing the economic profitability of loss of these peppers to the cost of an additional sensing point due to travel of the robotic arm of the harvester. The algorithm is validated on an additional dataset gathered according to protocols similar to the ones outlined in Chapter 3.2.
Next, given the algorithm for dynamic sensing that is capable of suggesting if a second viewpoint is needed and where it should be located, a **task planning algorithm** is developed in Chapter 6. This algorithm employs the traveling salesman paradigm (TSP) for optimization of harvesting and sensing sequencing to minimize the traveling time of the robotic harvester between the harvesting locations and the sensing locations. The algorithm evaluates the performance of different utility functions that measure the distance between locations. Additionally, it reviews different task planning strategies including those that sense first and then harvest all fruit or those optimizing the sensing location in the overall traveling sequence. To evaluate the algorithm a special technique for measuring exact location of peppers in the greenhouse is developed. Using this technique locations of peppers has been registered in the greenhouse. The developed algorithms were then evaluated based on the registered locations in laboratory and simulative conditions.
Finally, given the algorithms planning the harvesting sequence a method to **determine the approach direction** to each fruit before harvest is developed in Chapter 7. Using the pepper localization technique developed for the task planning algorithm in Chapter 6, peppers has been approached using different developed strategies in a repetitive manner in both laboratory and greenhouse conditions. In this chapter I also address the question of whether humans can detect the right approach direction by using questionnaires.

3.2 Data acquisition

3.2.1 Equipment and acquisition protocol

A Fanuc LR Mate 200iD/7L 6DOF robotic arm was equipped with three sensors and an illumination rig, simultaneously attached using the gripper described in Figure 3.1. The following sensors were used:

- iDS UI-5250RE RGB camera[1].
- Fotonic C70 RGB-D camera[2] based on time-of-flight methodology.

[1] https://en.ids-imaging.com/store/ui-5460re-poe.html
[2] Fotonic LTD has seized to exist, the camera is not available on the market. Similar here: https://acroname.com/sites/default/files/assets/fotonic_e-serien_2014_1_datasheet_0.pdf

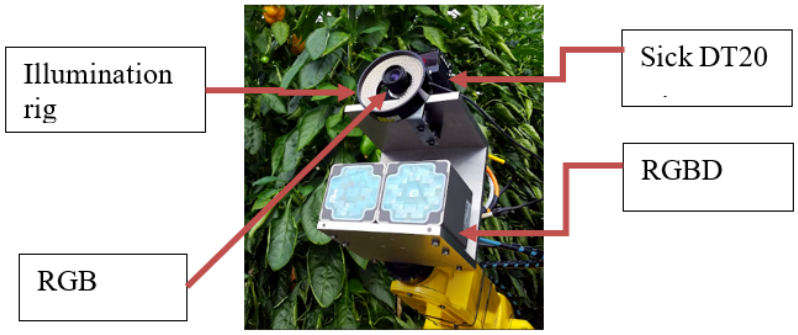


Figure 3.1: Robotic arm, sensor and lighting configuration.

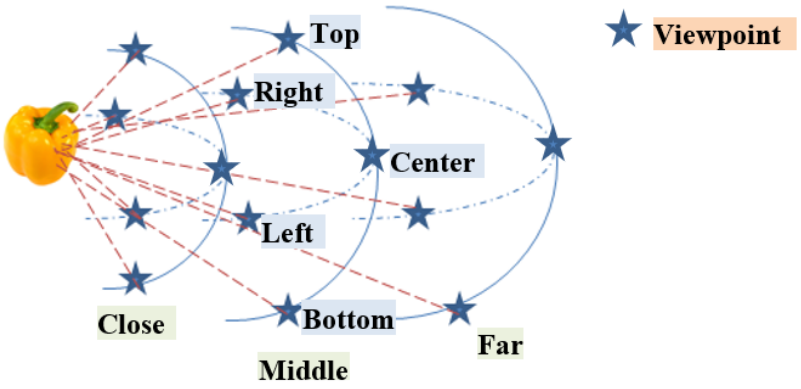


Figure 3.2: The 14 viewpoints from which image acquisition is performed.

- Sick DT20 Hi distance sensor[3].

The robotic arm was placed on a cart in an 80-cm wide isle (Figure 3.3). The iDS camera was placed in a way that the camera would in most cases face a mature pepper in its first position, and the height of the cart was manually readjusted accordingly. The robotic arm was configured to a planned sequence of 14 viewpoints in relation to the centered target, as shown in Figure 3.2. A user interface and an acquisition triggering software were written for image acquisition and image storage in C++. For the viewpoint evaluation described above, only the RGB images from the strobe light and natural light were used. An example of the acquired images is given in Figure 3.4. In general, in this thesis, the depth data produced by the Fotonic camera has not been applied. However, the data acquired in the process proposed as part of this thesis is available to the public via the sweeper databases as noted below.

3.2.2 Processing and labeling

Three human annotators were requested to perform labeling in which they were asked to perform segmentation of the image into pepper and background. The annotation was performed using GNU Image Manipulation Program (GIMP). Images from both natural light and strobe light were labeled.

[3]https://www.sick.com/ag/en/distance-sensors/displacement-measurement-sensors/dt20-hi/c/g176377

Figure 3.3: Data acquisition setup. Left: Isle in a sweet pepper commercial greenhouse. Right: Prof. Yael Edan, Dr. Boaz Arad and me during data acquisition.

Figure 3.4: Two example images taken from the same distance from two viewpoints. Left: an image taken from the right. Right: an image taken from the left.

3.3 Acquired datasets

The datasets acquired according to the defined protocol include the following:

1. **Viewpoints datasets**. 1312 images acquired from 10-14 viewpoints for 56 scenes were collected in commercial greenhouses. Three databases from different sweet-pepper varieties were collected from different growing seasons. Details on the data collected are described in Chapter 4, this data was used to evaluate algorithms presented in Chapters 4 and 7.
2. **Dynamic viewpoints dataset and 3D locations dataset**. 96 greenhouse images of 30 sweet peppers in a single research greenhouse from 12 viewpoints. Details in Chapter 5. The locations of each viewpoint and pepper were registered according to the protocol in Chapter 7 and used in Chapter 6.

An additional dataset used in this thesis was acquired according to the protocol in Chapter 7 used for determining the approach direction.

The first dataset from the *viewpoint datasets* has been made available for the public[4] as part of the Horizon 2020 SWEEPER data management plan. Two more datasets acquired according to a similar protocol to the one defined and are not used in this thesis will be also open to the public through the same media and through the submitted publication:

Boaz Arad, Polina Kurtser, Ehud Barnea, Ben Harel, Yael Edan, Ohad Ben-Shahar. "Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper harvesting robots". Submitted July 2018.

[4]Will be open to the public after Nov 2018, end of the Horizon 2020 sweeper project. Accessible here: http://icvl.cs.bgu.ac.il/lab_projects/agrovision/DB/Sweeper01/#/scene

4

Statistical models for fruit detectability: spatial and temporal analyses

> *Since I've written many of my books from a less-than-sympathetic viewpoint, I think that being able to see things from all sides is a useful talent.*
>
> — **Alex Flinn**
> (American writer of novels for young adults.)

- Published in: Kurtser, Polina & Yael Edan (2018). "Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers". Biosystems Engineering, pp. 272–289.
- **Research objective RO1:** Statistical evaluation of detectability from combination of viewpoints to provide insights into characteristics of best viewpoints.

Statistical models for fruit detectability were developed to provide insights into preferable variable configurations for better robotic harvesting performance.
The methodology includes several steps: definition of controllable and measurable variables, a data acquisition protocol design, data processing, definition of performance measures and statistical modeling procedures. Given the controllable and measurable variables, a data acquisition protocol is defined to allow adequate variation in the variables, and determine the dataset size to ensure significant statistical analyses. Performance measures are defined for each combination of controllable and measurable variables identified in the protocol. Descriptive statistics of the measures allow insights into preferable configurations of controllable variables given the measurable variables values. The statistical model is performed by back-elimination Poisson regression with a loglink function process. A spatial and temporal analysis is performed.
The methodology was applied to develop statistical models for sweet pepper (*Capsicum annuum*) detectability and revealed best viewpoints. 1312 images acquired from 10-14 viewpoints for 56 scenes were collected in commercial greenhouses, using an eye-in-hand configuration of a 6 DOF manipulator equipped with an RGB sensor and an illumination rig. Three databases from different sweet-pepper varieties were collected from different growing seasons.
Target detectability highly depends on the imaging acquisition distance and the sensing system tilt. A minimum of 12 training scenes are necessary to discover the statistically significant spatial variables. Better prediction was achieved at the beginning of the season with slightly better prediction achieved in a temporal split of training and testing sets.

4.1 Introduction

Despite intensive R&D on harvesting robots, to date no commercial harvesting robot exists [Bac et al., 2014b]. One major limitation of current developments is the low detection rates of around 87% [Bac et al., 2014b] caused by the complex and highly variable agricultural environment.

The most common placement of the vision sensor in robotic harvesters till today in an eye-in-hand configuration [Bac et al., 2014b]. Current agricultural robotics detection algorithms usually use a single preset viewpoint [McCool et al., 2016; Sa et al., 2016; Bac et al., 2014b], however since single viewpoint visibility is limited [Hemming et al., 2014b; Bulanon et al., 2009] multiple viewpoints are necessary to improved detection rates. Since cycle times are critical [Bac et al., 2014b] it is important to direct the robot to a minimum number of viewpoints that can provide maximal detectability.

The search for an optimal viewpoint has been extensively investigated in many fields [Foix Salmerón et al., 2011; Fleishman et al., 2000; Reed & Allen, 2000; Maver & Bajcsy, 1993]. However, the high scene variability in a harvesting application, which is inherent to the biological nature of the scene, reduces the ability to calculate a-priori the best viewpoints due to the very limited geometric information about the scene. The complexity of the fruit detection task is due to the unstructured and dynamic nature of both the objects and the environment [McCool et al., 2016; Sa et al., 2016; Gongal et al., 2015; Kapach et al., 2012]: fruits have a high inherent variability in size, shape, texture, and location; in addition, occlusion and variable illumination conditions significantly influence the detection performance. This research aims to determine the dependency of detectability on the chosen viewpoint, and to analyze the temporal and spatial relations between consecutive scenes. The proposed methodology for developing statistical prediction models of fruit detectability for vision based robotic harvesters enables to prove correlations between controllable and measurable variables and the detectability performance measures, providing insights into preferable configurations for better robotic harvesting performance.

The chapter starts with a literature survey of common practices in detectability and visibility research. Section 4.3 outlines a methodology for viewpoint detectability modeling and definition of performance measures. It aims to present the minimum size of training sets for which the suggested controllable variables are still found to be significant and therefore will lead to a correct sensing plan that will increase detectability. Section 4.4 presents the results of application of the proposed methods on a case study database of *Capsicum annuum* (sweet peppers). Section 4.5 outlines the drawn conclusions and Section 4.6 summarizes the relevant research questions and answers.

4.2 Background

4.2.1 Detectability and visibility

Current agricultural robotic detection algorithms aim to maximize the true positive rate and minimize the true negative rate [Vitzrabin & Edan, 2016b]. The standard procedure applied in the computer vision community for defining true positive and true negative rates is labeling images [Russell et al., 2008]. The labeling process includes either a bounding box or pixel-wise labeling resulting in segmented image into areas representing the targets and the background. This is performed by annotators reviewing the images for unlimited amount of time [Deng et al., 2009; Russell et al., 2008]. In some cases the annotators are requested to classify the targets annotations into fully revealed targets, partially occluded targets and truncated targets [Geiger et al., 2012]. The labels are often used both as a training set of supervised learning algorithms as well as a benchmark for the detection algorithms.

However, the annotation of a single image is not sufficient for object detectability within a scene. The number of targets within the scene is defined as the ground truth. The common procedure for obtaining detectability ground truth is placing several sensors that simultaneously sense the same scene and then combining and annotating the joint number of targets in the scene [Dollar et al., 2012; Russell et al., 2008]. By doing so, a benchmark of joint detectability using vision sensors is generated. The placement of the sensors in a way that covers the whole operational area is critical for proper ground truth acquisition, providing detection of all targets relevant to the robotic task (e.g., reachable by the robotic manipulator, to be avoided by an autonomous vehicle).

An alternative type of ground truth, that has been used for agricultural applications [Bac et al., 2014b; Hemming et al., 2014b], and some SLAM applications [Blanco et al., 2009] is the visibility ground truth. Visibility in robotic harvesting is defined as "The visible part of a fruit in an image expressed as a percentage of total fruit area which would be seen in an image without occlusion" [Hemming et al., 2014b]. To obtain the visibility ground truth an in-field human observer manually counts the actual number of targets in the scene. The visibility analysis compares between targets manually labeled from the vision system to the actual number of targets present within the scene. This allows a visibility benchmark given sensors capable of detecting targets occluded by other objects in the scene. While it is an important analysis for evaluating the final performance of a robotic harvester, it does not provide a true benchmark for vision-based robots that cannot "see-through" obstacles.

Furthermore, it is limited in the number of data that can be analyzed since it requires accurate counting of all fruit in the analyzed scenes. Reported numbers should be separated into two – fruits that grow on trees with high density of targets on a single plant (e.g., citrus or apple trees) and fruits that grow on stems, where the density of fruits in a scene is low (e.g., peppers, cucumbers, eggplants). The latest research groups in the field report a 40-150 targets for apple trees [De-An et al., 2011; Sites & Delwiche, 1988], 150-900 targets for citrus trees [Lee & Rosa, 2006; Plebe & Grasso, 2001]; although these numbers are relatively high they do not enable systematic statistical modeling since they represent data from 1-2 trees.

The quantity of ground truth drops significantly for low density crops with 60-200 targets for cucumbers [Tang et al., 2009; van Henten et al., 2002] and 40 targets for eggplants [Hayashi et al., 2002].
To develop a statistical modeling methodology for determining the optimal viewpoints and the fruit detectability a dynamic protocol that defines the changes in environmental and hardware conditions must be repeated a sufficient number of times for each condition combination. This requires a large database. In order for the protocol to be performed autonomously, without human intervention, the ground truth gathering should be automated. Current algorithms' detection rates are low and hence, do not allow automatic detection to be treated as ground truth. Therefore, the common practice for ground truth acquisition protocols is based on manual processing of images acquired. As a result the models presented in this research focus on detectability modelling. The detectability ground truth is gathered by manual evaluation of all viewpoints from a given scene to determine the overall number of targets detectable in the scene.
Most robotic harvesters are equipped with a vision sensor in an eye-in-hand configuration [Bac et al., 2014b]. Some harvesters include additional sensors such as RGB-D cameras, or dual cameras used both on the robotic manipulator and the robotic cart statically placed within the greenhouse lane [Barth et al., 2016; Bontsema et al., 2015; Bac et al., 2014b]. Due to the dynamic nature of the agricultural scene sensor fusion is highly unreliable for localization of the harvesting targets [Barth et al., 2016]. Additional sensors also increase the overall complexity and price of the robotic harvester. Therefore, additional static sensors are uncommon and usually vision based robotic harvesting relies on a single sensor placed on the robotic manipulator [Barth et al., 2016; Bac et al., 2014b]. A subset of viewpoints is chosen to perform detection at each scene. The viewpoints are reached by the robotic manipulator equipped with a camera. At each viewpoint an image is captured followed by its analysis. Each additional viewpoint requires additional travel time by the manipulator and additional computational time of the image processing. Therefore, the number of viewpoints should be minimized.

4.2.2 Viewpoints analysis

A viewpoint can be described as a set of geometric and optical sensing parameters [Tarabanis et al., 1995]. The geometric parameters are the six degrees of freedom representation of the position and orientation of the sensor while the optical parameters depend on the lens setting (e.g., focal length, field of view). The process of a best viewpoint search mostly depends on how much a-priori geometrical information about the scene is available [Maver & Bajcsy, 1993].
The search for an optimal viewpoint has been extensively investigated in many fields, such as computational geometry [Vázquez et al., 2001; Maver & Bajcsy, 1993], graph drawing [Vázquez et al., 2001], and multi-camera stereo vision [Zabulis & Daniilidis, 2004]. In robotics applications optimal viewpoint analyses have been investigated for motion planning [Vázquez et al., 2001], grasping [Krainin et al., 2010], and medical intervention planning [Mühler et al., 2007]. Most research aims to calculate a global optimal placement of sensors

(usually a camera), to maximize detection, mutual information, or some other cost function appropriate in the given application. These methods often rely on generating image-based models from scenes with known [Fleishman et al., 2000], or unknown [Reed & Allen, 2000] geometry, for RGB [Fleishman et al., 2000] or TOF [Foix et al., 2015] cameras.
In agricultural robotics most work to date reported on object detectability [Kamilaris & Prenafeta-Boldú, 2018; Bac et al., 2014b; Kapach et al., 2012; Bulanon et al., 2009], and relied on detection from a single viewpoint. A next best viewpoint algorithm for leaf segmentation on data acquired from pots in the lab has been proposed [Foix et al., 2015; Foix Salmerón et al., 2011]. Manual analyses of 30 sweet pepper plants revealed that only 40-60% of the fruit present in the field of view are visible from a single viewpoint, by a human observer [Hemming et al., 2014b]. Similar research on citrus harvesting [Bulanon et al., 2009] for a total number of 5 trees reported a combination of viewpoints raises the detectability from 40-70% to 80-90%. Hence, multiple viewpoints are necessary to overcome the detection rate bottleneck limiting harvesting performance.
The number of necessary viewpoints depends on the expected system variability. The variability is caused by the object, environment and robotic variability. The object variability is a characteristic of the biological inherent variability, in addition to the variation caused by the different growing and environmental conditions, resulting in differences in size, color, shape, location and texture of the targets. The environment variability includes unstructured obstacle locations (e.g., leaves, branches) and changing lighting conditions (e.g., daylight, night light, bright sun, shadows), which depend on the time and location and directly affect the detection performance. Hence, the quality of target detection often relies on the presence of artificial illumination [Vitzrabin & Edan, 2016a; Tarabanis et al., 1995]. The specific robotic system modules used (the sensors, illumination and manipulator design, including degrees of freedom, dimensions and controls) also affect the detection performance. While the object variability and environment variability may be supported by botanical models to some extent, their combination with the unpredictable and dynamic lighting conditions (e.g., sun direction, clouds), the growing conditions (e.g., climate, weather, cultivar, density of plants) and the robotic system (specific sensors, illumination, robot) makes it difficult, if not impossible, to develop an analytical model to predict fruit detectability. When analyzing detectability it is important to measure also the degree of exposure of the peppers. Fully revealed peppers are easier to be detected for current detection algorithms [Hemming et al., 2014b], and fully revealed peppers are easier to be localized for grasping [Eizicovits et al., 2016]. Therefore, each viewpoint should also receive a weighted score that gives more weight to views including fully revealed targets than to views with partially occluded or truncated targets.
For efficient robotic harvesting more than 90% visibility is necessary [Blackmore et al., 2002]. Since cycle times are critical [Bac et al., 2014b] it is important to direct the robot to a minimum number of viewpoints that can provide maximal detectability. This research is directed by the concept that by developing a fruit detectability statistical model, the detectability for a given set of conditions can be predicted enabling to select the best-fit viewpoints for the particular environmental conditions and hardware. By determining the minimal training set size used

for prediction of detectability, one can define how often the detectability prediction model should be re-trained given changes in the environment (e.g., growing practices, season/time of day) and operating conditions (e.g., robot, camera, illumination). A larger training set will require less often re-training however, will less fit the dynamic changes. To the best of our knowledge, this is the first attempt to statistically model detectability for harvesting robotics. The determination of the best-fit viewpoints can lead to improved robotic performance.

4.3 Statistical model development methodology

4.3.1 Overview

The statistical models for fruit detectability include the following steps (Figure 4.1):

1. **Definition of controllable and measurable variables**. These are the commutative variables controlling the viewpoint (such as pose and orientation), the illumination (e.g., intensity, direction) and describing the environment (such as cultivar and growing practices, time along the growing season).
2. **Data acquisition protocol design**. This is the ground truth viewpoint information acquisition and experimental design phase enabling data acquisition in a range of values of the defined variables. The experimental protocol determines the dataset size, as well as the combinations of variable values.
3. **Data processing**. Including labeling and tagging of the targets by a human observer who marks a ROI of the target (e.g., fruit). The fruit may be additionally classified into fully visible, partially occluded and truncated fruit as common in the computer vision community [Geiger et al., 2012].
4. **Definition of performance measures and statistical modeling**. The performance measure of a good viewpoint is defined in a commutative manner. It provides an overall descriptive statistical evaluation of the current detectability and weighted viewpoint score, given the measured variables, and evaluates the prediction capabilities, given new values of the variables.

4.3.2 Controllable and measurable variables

Each additional variation of environment, object and robotic features carries additional observations. Therefore, to generate a detectability model, acquisition of large datasets in predefined conditions are required. This is often challenging due to the harsh outdoor conditions. Therefore, this research aims to determine a subset of controllable and measurable variables that can be automatically acquired by a robotic harvester to reduce the volume of acquired data. A statistical model is generated only for those variables and in a back elimination process the suggestions for best variable combinations are produced. In this case the data quantity required is still significant, but is obtainable using an automatic acquisition protocol.

The controllable spatial parameters aim to define the relation between the geometrical

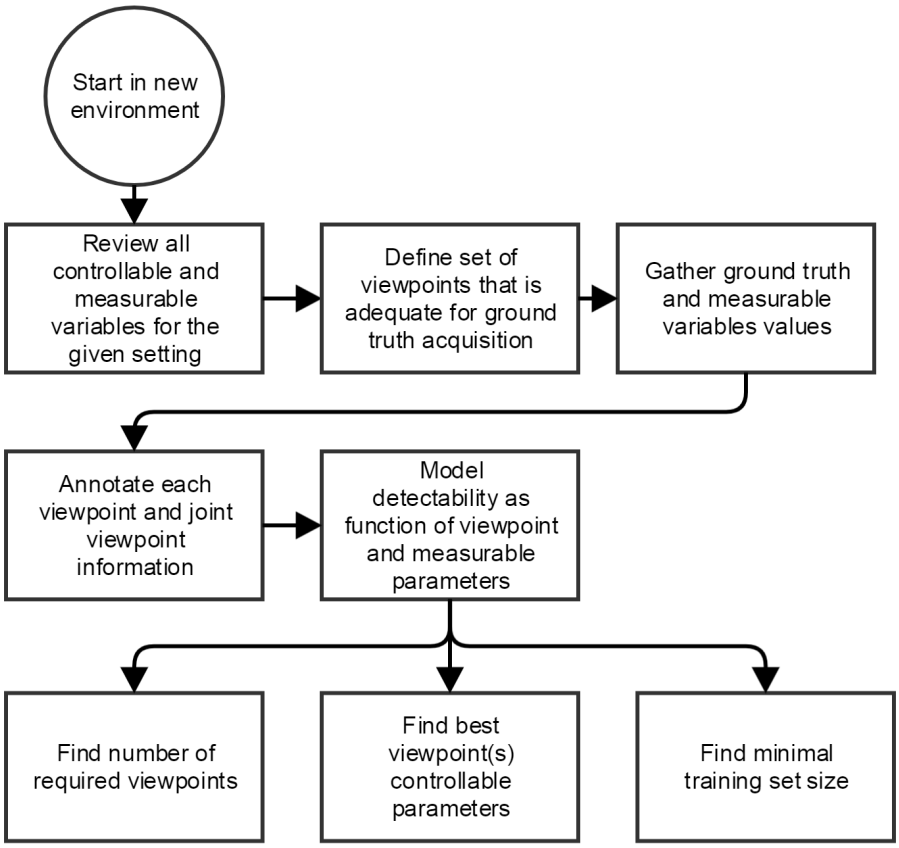


Figure 4.1: Flowchart of overall statistical modeling of detectability.

placement of the sensor compared to the fruits position and the greenhouse lane. This includes distance, tilt and azimuth angle that are derived from both the sensors' geometric parameters (e.g., position and orientation) as well as the position of the robot within the greenhouse lane, the distance between lanes, the height of the cart on which the robot is placed, physical size of the robot, and the height of the plants.

Machine vision in outdoor conditions must cope with highly varying illumination conditions [Tian & Slaughter, 1998]. Hence, it is important to ensure the dataset includes a large variability in illumination; this is achieved by acquiring images at different times along the day, at different locations within the greenhouse and at different angles towards the sun.

Although natural illumination intensity can be measured using an external illumination sensor since the values are subject to extreme changes in a matter of split seconds, addressing the natural light magnitude as an accurate measurable parameter was excluded. Therefore, this is an uncontrollable variable. Some methods aim to cope with natural illumination by using artificial illumination [Sun et al., 2007; Jimenez et al., 2000]. Therefore, the presence of artificial illumination is defined as a controllable variable.

The environmental parameters are defined as measurable parameters. These are the most complicated parameters to quantify due to the aforementioned high variability which are not always known in advance. The environmental parameters that highly challenges the detection are those affecting the expected occlusion. This includes the orientation of the lanes

and the side of the lane on which the harvesting is performed [Rylski & Spigelman, 1986], the fruit variety, the plant hardiness zone which determines if a plant is likely to thrive at a location [Daly et al., 2012], time since planting and time since last harvest.
To focus the question on the choice of the correct sensing point given a static hardware setting, for the purpose of this thesis, all hardware related parameters including optical parameters (e.g., lens configuration), robotic parameters (e.g., DOF, length of robotic arm), and illumination parameters (e.g., flash illumination) are assumed to be measurable and not controllable. This allows repeatability in greenhouse conditions, minimizing the amount of physical changes performed in the field and therefore minimizing the chance of human error and hardware fault in the acquisition protocol.
Given an eye-in-hand sensing configuration, the controllable variables are defined as follows:

- **Spatial variables**. Variables describing the spatial orientation of the sensory system. A viewpoint is defined by three controllable spatial variables (Figure 4.2) – distance to fruit, tilt angle, and azimuth angle.
- **Camera variables**. Variables describing the sensing equipment (e.g., lens configuration, resolution and aperture).
- **Illumination variables**. Variables describing the presence of artificial illumination.

In addition to the controllable variables, measurable variables that influence the target detectability are defined as measurable environmental variables. These variables are environmental dependent variables that describe the scene in which the robot operates. The variables include: number of weeks since beginning of harvesting season, days since last harvesting and direction of the natural light (e.g, the orientation of the lanes in terms of north/west and the side of the lane being harvested - Figure 4.3).
At the beginning of the growing season, the scene is less occluded since it has less leaves, allowing better target detectability. The number of days since last harvesting affects the number of targets to be detected. The direction of the natural light can create overexposure of the images acquired if directed towards the cameras and, therefore, creates a challenge for detection, both for a human observer and for autonomous detection algorithms (Figure 4.4).

4.3.3 Data acquisition protocol design

The acquisition protocol is defined to ensure images are acquired in a wide range of possible controllable variables (i.e., different viewpoints and different illumination conditions) and environmental variables (i.e., different greenhouses, different times along the season and different orientations to the sun).
The databases should be acquired according to the protocols described in Figure 4.5, in different environmental and acquisition conditions, to provide a wide range of values for the measurable and controllable variables.
The viewpoints must be defined so as to cover the workable area of the robotic harvester, under the assumption of obstacles randomly oriented and occluding the fruits. A robotic harvester operates in the greenhouse by picking fruits from each side of the growing lane,

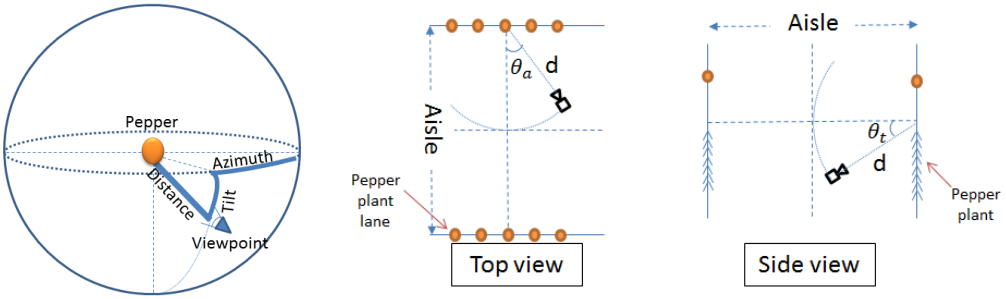


Figure 4.2: Definition of a viewpoint by distance and two angles – tilt and azimuth. Left: the definitions; Middle and Right: estimations by defining the distance to be the distance to the isle and the azimuth and tilt angle to be θ_a and θ_t accordingly as seen from a top and side views of the greenhouse lane.

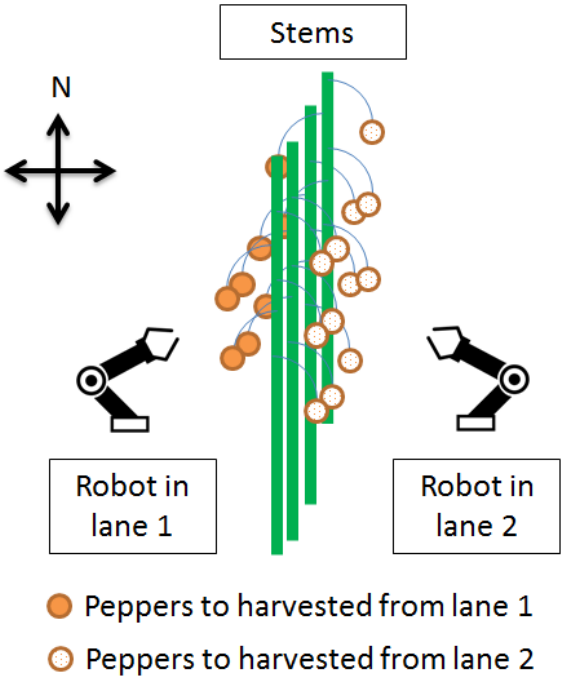


Figure 4.3: Peppers to be harvested from each side of the stems are those between the robot arm and the stems. Given a stem oriented south-north the peppers to be harvested from lane 1 are facing west and the ones to be harvested in lane 2 are facing east.

Figure 4.4: Images taken in different seasonal conditions. Left: beginning of season, 3 days since last harvesting; Middle: end of season, harvesting performed 2 days prior acquisition; Right: challenging lighting conditions when images taken against the light.

similar to the operation of manual workers in the field. Attempting to harvest a plant only from one side of the lane, by reaching fruits hanging on the opposite side of the stem, is in risk of hurting the stems which are a viable part of the plant causing severe financial lost to the greenhouse [Bac, 2015; Bac et al., 2014b]. Therefore, fruits hanging on the opposite side of the stem, where the stem is in between the camera and the fruit should be harvested from the opposite side of the stem in the next lane (Figure 4.3). Algorithms have been developed to avoid the stem [Bac et al., 2016] and perform harvesting of such fruits, but they do not justify the robotic operation on only one side of the lane. Furthermore, the further the fruit is hanging from the robotic arm the more obstacles are present both for detection and for reaching the fruit without hurting the plant. Fruits that are not visible from either side of the lane are to be disregarded and cannot be harvested.

As a result, for obtaining ground truth information, the viewpoints must be selected to cover half a cylinder of radius R_S around the stem (Figure 4.6). R_s is to be defined by the expected size of the fruit peduncles of the specific fruit to be harvested. The robotic manipulator is capable of covering a sphere of radius R_r. The intersection between the robotic approachable sphere and the cylinder of interest generates a reachable area by the robot within the vegetation.

Given a sensor with vertical and horizontal angles of view of θ_{CV} and θ_{CH} the area within the robot reachable area within the vegetation is defined as presented in Figure 4.6, assuming no occlusions are present. As a result, the combination of viewpoints required for ground truth acquisition is defined in a way they will cover the robot reachable area within the vegetation (Figure 4.6).

4.3.4 Data processing

The collected RGB images data should be tagged for ground truth collection. The tagging process includes marking of a rectangular ROI around the object by an annotator. The ground truth collection process is according to the detectability ground truth collection process described in Section 4.2.1. Additionally, the annotator's performance should be measured by providing the distribution of size of the pepper tagged as described in Section 4.3.5. The applied protocol is described in details in Section 4.4.2.

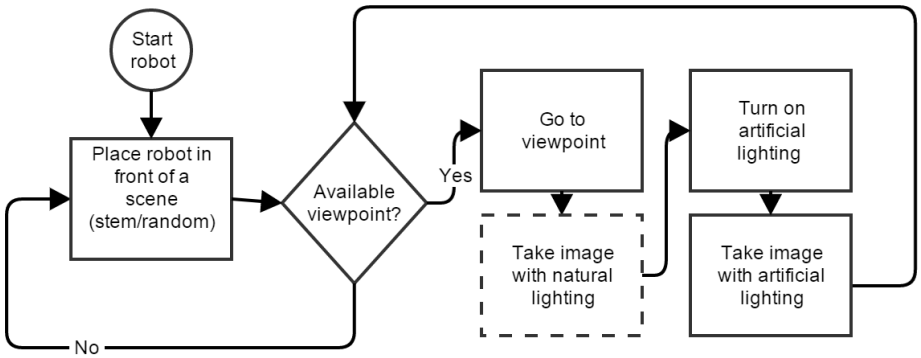


Figure 4.5: Acquisition protocol.

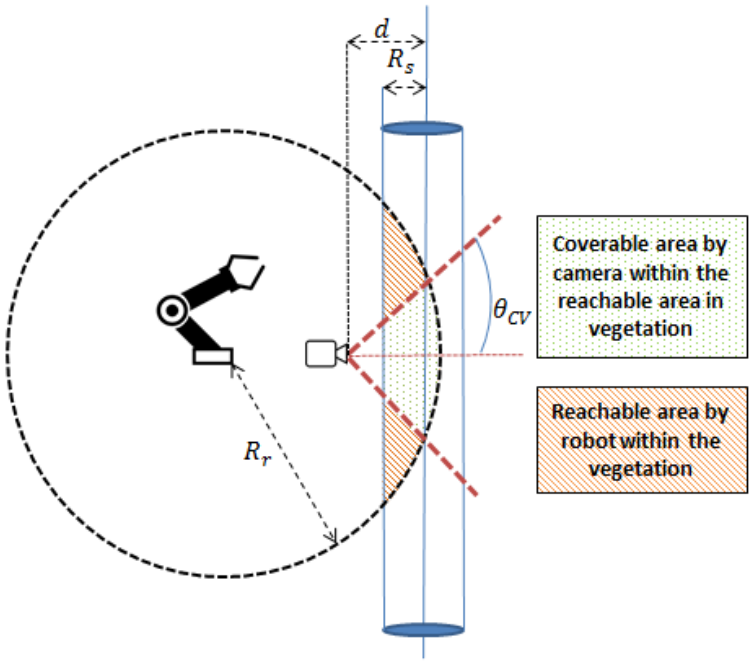


Figure 4.6: Viewpoints covering area with no occlusion present.

4.3.5 Definition of performance measures and statistical modeling

Predictive statistical models versus correlation exploration and descriptive statistics

The detectability and visibility research for harvesting robotics to date [Hemming et al., 2014b; Bulanon et al., 2009] focused on finding a set of controllable parameters under the presence of measurable parameters that will be positively correlated to the visibility and detectability rates. This allows insights into what controllable variables values are preferable for better detection and visibility. This kind of descriptive statistics and correlation exploration, define the possible variables implicating detectability in the setting described.

However, the ability to generalize the model to new observations from the same population,

and to other populations [van Houwelingen & Le Cessie, 1990] is another important aspect that must be addressed. After defining a set of parameters that positively correlates with detectability and visibility a prediction model of detectability for a new scene is viable. This enables to generate a sensing plan to be executed by the robot prior to harvesting. As aforementioned, the generation of an analytical model is highly unlikely in the highly variable setting, where all parameters cannot be determined a-priori and even when defined cannot always be tracked. Furthermore, no attempt to prove causation relations [Wright, 1921] between the controllable and measurable parameters and the predicted detectability is needed for the robot to perform adequately.

The model aims to provide an empirical prediction model based on data gathered in similar environmental and robotic conditions to the one the harvester will be operating in. This allows homogeneity in the unmeasurable environmental parameters that creates more adequate combination of controllable variables to the described setting. The prediction model is evaluated for its prediction capabilities using goodness of fit measures [van Houwelingen & Le Cessie, 1990; Willmott, 1981].

The modeling methodology can be performed repetitively before each harvesting session for the specific data. A harvesting session can be defined as a greenhouse lane, day of the week, growing month or any other interval that supplies additional variability for which a previously trained model is not valid any longer. The modeling protocol allows to define the ranges of controllable parameters values that are most suitable for the current setting, and the evaluation of significant and insignificant variables. An additional requirement for defining a repetitive detectability model is the definition of the training set, as described below.

Sample size

In order to find statistically significant variables, the null hypothesis of absence of effects must be rejected. The power of the reject depends on the number of samples [Murphy et al., 2014], where more samples enlarge the chances to find statistically significant variables. Obtaining detectability ground truth information requires annotators reviewing large number of images for each training scene which is a tedious task and is a time consuming operation limiting the amount of data to be analyzed.

Therefore, an important decision is the minimal sample size necessary to define a prediction model for a given set of variables. This question is addressed by modeling the detectability based on training datasets of varying sizes. The largest training dataset defined the variables significance for detectability prediction in the given environment. The minimal training set is defined as the minimal training set size for which the variables that were found significant in the largest dataset are still significant.

To develop the model, the acquired dataset is usually split between training and testing data sets [Sa et al., 2016; Guyon, 1997]. The spatial-temporal relations between consecutive scenes can influence the model development and hence this must be considered in the splitting of data. In a random training-testing split (Figure 4.7) random scenes are used for training of the model and testing of its goodness of fit. Since additional variations in objects and environment are expected due to temporal and spatial effects, where scenes acquired chronologically in space or time have more common features than random scenes

	Scenes											
Random split 80:20 training set size:3	1	2	3	3	2	3	1	2	1	1	3	2
Random split 80:20 training set size:2	1	4	1	3	2	2	4	3	4	1	2	3
Temporal split 80:20 training set size:3	1	1	1	1	2	2	2	2	3	3	3	3
Temporal split 80:20 training set size:2	1	1	1	2	2	2	3	3	3	4	4	4

Training scene of set X	X
Testing scene of set X	X

Figure 4.7: Example of a set splitting into training and test sets.

taken within the same experiment, the training and testing scenes can be considered to be chronologically split as opposed to a random split. In this case of temporal split the detectability prediction model trained on a set of consecutive scenes is applied to test scenes chronologically placed after the training set (Figure 4.7). The random splits expected to yield a more general detectability prediction model that is less fitted for each individual scene while the temporal split is expected to yield more accurate prediction results.

Descriptive measures

Given the detected number of targets in the image and the ground truth information collected for the overall number of targets available in the scene, two measures were defined, as described in Equations 4.1-4.2. The proportion of visible targets P_i is defined in Equation 4.1.

$$P_i = \frac{T_i}{T} \tag{4.1}$$

where T_i is the number of targets detected from viewpoint i and T is the joint number of targets seen from all viewpoints taken at the scene. The proportion of visible targets from multiple viewpoints P_{ij} is defined in Equation 4.2.

$$P_{ij} = \frac{T_{ij}}{T} \tag{4.2}$$

where T_{ij} is the joint number of targets seen from viewpoints i and j.
The detected targets can be fully visible, partially occluded or truncated (Figure 4.8). The weighted score of a viewpoint, as function of detected target type k (**F**ully visible, Partially-**O**ccluded or **T**runcated) is defined according to Equation 4.3.

$$S_i = \sum \frac{W_k T_{i(k)}}{T}; \forall k = \begin{cases} F & Fully-visible \\ O & Partially-occluded \\ T & Truncated \end{cases} \tag{4.3}$$

where w_k is the relative weight given to each target of type $k \in (F,O,T)$, and $T_{i(k)}$ is the number of targets detection from viewpoint i of type k.
The following weight vectors are defined:

- **Detectability weights**. In this case the weight of all detected peppers are equal. Important to note that given equal weights for each target type ($w = \vec{1}$) the score results in $P_i = S_i$.

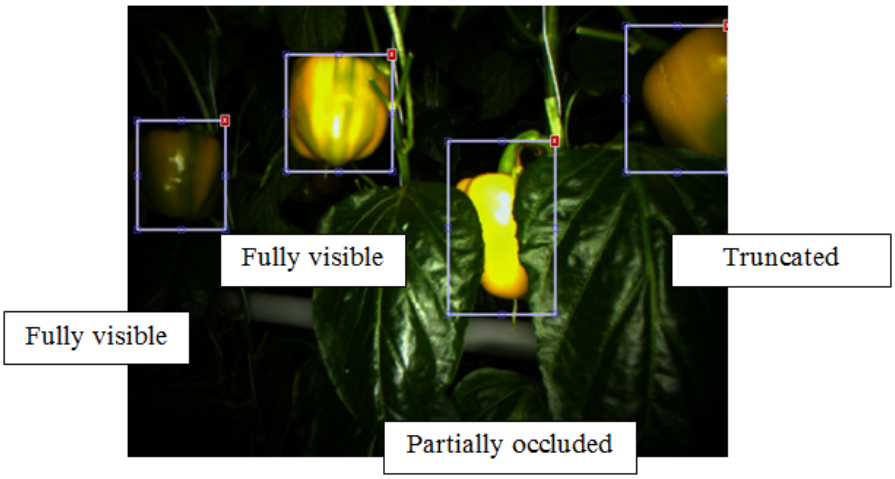


Figure 4.8: Example of a tagged image and the classification of the marked peppers into ROIs of 3 types: fully visible, partially occluded, and truncated.

- **Relative weight**. Given the assumption that occluded peppers are more difficult to detect than truncated or full visible peppers a weight of 0.5 is given for occluded peppers, 0.7 for truncated peppers and 1 for fully visible peppers.

Additionally, to evaluate the minimal size of a detectable pepper by an annotator the bounding box area A_l is measured as the number of pixels of the marked bounding box for pepper l. Then the relative bounding box area RA_l is defined according to Equation 4.4.

$$RA_l = 100\frac{A_l}{M * N} \tag{4.4}$$

Where M and N are the height and width of the image respectively.

Detectability modeling under spatial and temporal viewpoints

A back-elimination Poisson regression with a log link function [Winkelmann & Zimmermann, 1995] modeling technique was performed to model the relation between the controllable parameters and the detectability of the targets in given measurable environmental conditions. This technique, especially suitable for count-data such as the number of detected fruit in an image, enables to define a short list of statistically significant controllable variables. Modeling the number of detected peppers in an image as a function of the controllable variables provides additional insights. The outputs include:

- **Statistically significant variables**. These enable distinction between differences in detection rate correlated to the controllable parameters and empirical random error.
- **Regression weights**. These enable analysis of the correlation between the statistically significant controllable variables and the number of detected peppers in terms of positive and negative influence (positive weights and negative weights).
- **Statistically significant interaction variables**. By including the interaction variables, an analysis of the joint relation of the controllable variables on detectability prediction is performed.

- **Prediction model**. Prediction of the expected number of peppers, given new values of the significant controllable parameters.

While the statistically significant variables (independent or interactions between them) and regression weights are investigated to research the relation between the variables and detectability, the prediction model aims to explore the ability of the given variables to predict the detectability for a given viewpoint. Each tagged dataset was separated into two sets: (i) a training set that includes a subset of scenes used for learning the model of the number of targets as a function of the viewpoints variables, and (ii) a test set to evaluate the prediction model, given a new scene.

Since some variations of the object and the environment are not random, consecutive scenes are expected to have common features. To explore the temporal relations, as opposed to general spatial relations, within the data, the datasets were split using two different methods: ***random split***, and ***temporal split***. The random split divides the entire dataset into random training sets and random test set according to the defined train-test ratio and sample size (Figure 4.7). The temporal split defines the number of consecutive scenes involved in the training set and the number of consecutive scenes involved in the test set, where the test set is chronological to the training set (Figure 4.7). Consecutive temporal split training sets generated are overlapping.

To evaluate the sensitivity of the model to the number of scenes included for the training set, and to derive the minimal number of scenes required for exploring the relation between the detectability and the viewpoint spatial variables, the following ratios were considered: 90:10,80:20,70:30,60:40,50:50,40:60,30:70,20:80,10:90. For each split dataset the following measures are calculated:

- **Minimum training set size**. The significance of the variables found statistically significant for the full database is re-evaluated for the smaller training subset to define the minimal training set for which the found spatial variables remain statistically significant.
- **Root mean square error (RMSE)**. A goodness of fit measure of the prediction model, defined in Equation 4.5:

$$RMSE = \sqrt{\sum \frac{(T_i - \widehat{T}_i)^2}{N}} \tag{4.5}$$

  where T_i is the actual number of detected peppers, $\widehat{T}_i$ is the predicted number of detected peppers and N is the number of scenes in the test set used for the evaluation.
- **Willmott's index of agreement (IOA)**. A goodness of fit measure [Willmott, 1981] for the prediction model defined in Equation 4.6.

$$IOA = 1 - \frac{\sum (T - \widehat{T}_i)^2}{\sum (|\widehat{T}_i - \bar{T}| + |T_i - \bar{T}|)^2} \tag{4.6}$$

  where $\bar{T}$ is the average number of detected peppers in a viewpoint.

4.4 Detectability analysis for a case study of sweet peppers

4.4.1 Databases

A dataset of 1312 images of sweet peppers (*Capsicum annuum L.*) with a total of 798 sweet peppers annotations were acquired in three different greenhouses in different seasons (beginning, middle and end of the season) using two different acquisition methods (manual and autonomous) and two illumination types (artificial and natural). More details about the databases are given in Table 4.1 while the list of used viewpoints is defined in Table 4.2. The viewpoints are determined according to the guidelines defined in Section 4.3.3, given a 2-4 cm length of the peduncle R_S for sweet peppers [Eizicovits et al., 2016]. The acquisition was performed as follows:

1. **Fully autonomous, mid-season database (DB#1)**. The database was acquired in a commercial greenhouse in Ijsselmuiden, Netherlands with growing lanes oriented approximately southwest-northeast, using a 6 degree of freedom manipulator (Fanuc LR Mate 200iD/7L), equipped with an iDS Ui-5250RE RGB camera with a Lensagon CMFA0420ND 4.166mm lens set at minimal aperture to automatically acquire images from 14 viewpoints defined with both artificial and non-artificial illumination conditions [Kurtser et al., 2016].
2. **Manually acquired, end of season database (DB#2)**. The database was acquired manually in a commercial greenhouse in Kamehin, Israel, growing lanes oriented approximately north-south, using the same RGB camera with a Tamron M118FM12 lens with a focal length of 12mm and minimal aperture mounted on a 3 degree of freedom tripod, both in artificial and non-artificial illumination conditions.
3. **Fully autonomous, beginning of season database (DB#3)**. The database was acquired in a research greenhouse at Sint-Katelijne-Waver, Belgium with growing lanes oriented approximately southwest northeast and artificial growing lights conditions, using a 6 degree of freedom manipulator (Fanuc LR Mate 200iD), equipped with the same RGB camera with a Lensagon CMFA0420ND 4.166mm lens set at minimal aperture to automatically acquire images from 12 viewpoints with artificial illumination conditions.

Two acquisition protocols were applied for image acquisition (Figure 4.5): a centering approach and random acquisition approach. In the centering approach, applied to DB#1 and DB#3, the sensory system was manually centered in front of a pepper or a cluster of peppers, and then the acquisition was performed. DB#1 includes several random positions and rows within the greenhouse, while DB#3 is a smaller set acquired from a single row on two sides with scenes spatially ordered one after the other. DB#3 did not include any natural lighting conditions. In the random acquisition approach, applied to DB#2, the system was manually set to a given height and then the acquisition was performed by manually moving the sensory system along the lane; images were acquired at a constant distance of 1m apart.

Table 4.1: Databases description

Feature	DB#1	DB#2	DB#3
No of scenes	32	16	8
Illumination conditions	Artificial/Natural	Artificial/Natural	Artificial
Growing lights	Natural	Natural	Artificial
Lanes orientation	SW-NE	N-S	SW-NE
Greenhouse	Netherlands	Israel	Belgium
Weeks since season start	22	45	3
Days since last harvest	3	2	3
Acquisition mode	Autonomous	Manual	Autonomous
Resolution	1280X960	1600X1200	800X600
Focal length	4mm	12mm	4mm

Table 4.2: Viewpoint description for each database.

DB	VP	Distance (mm)	Tilt (degrees)	Azimuth (degrees)	VP	Distance (mm)	Tilt (degrees)	Azimuth (degrees)
DB#1	1	450	0	0	8	300	22	0
	2	450	0	45	9	300	-22	0
	3	450	0	-45	10	350	0	0
	4	450	22	0	11	350	0	45
	5	300	0	0	12	350	0	-45
	6	300	0	45	13	350	22	0
	7	300	0	-45	14	350	-22	0
DB#2	1	700	0	0	6	400	0	0
	2	700	0	45	7	400	0	45
	3	700	0	-45	8	400	0	-45
	4	700	30	0	9	400	30	0
	5	700	-30	0	10	400	-30	0
DB#3	1	19	20	0	7	10	20	0
	2	19	-20	0	8	9	0	0
	3	8	-20	0	9	7	20	20
	4	19	10	-50	10	7	0	20
	5	10	10	-50	11	19	20	20
	6	10	0	-50	12	17	0	0

4.4.2 Target identification - image processing, annotation and ground truth measurements

The acquired images were processed by window based manual tagging, using the $Matlab^{LTD}$ Training Image Labeler tool (Computer Vision System Toolbox). The annotator was requested to draw a bounding box around all visible ripe peppers. A ripe pepper was defined as a pepper with a minimal coloration of 20% of the visible area evaluated visually by the annotator. In case of truncated or occluded peppers, the annotator was asked to draw a ROI around the visible area only. The annotator continues until satisfied with the results. The result of the tagging is a set of peppers ROIs, as shown in Figure 4.9. The annotator was also requested to classify each tagged pepper in each image of DB#3 into one of three categories: (i) fully visible; (ii) partially occluded; (iii) truncated. The annotator was directed to mark fully visible peppers if they are not occluded or truncated by more than 10% of the area. In case peppers were both occluded and truncated, the peppers were classified as occluded. No specifications of the occlusion type (stems, leaves, other peppers, foreign object in the images) were noted. The output of the labeling process is a set of image coordinates representing the tagged peppers. The tagging process is then translated into a dataset, defining the number of targets tagged for each image.

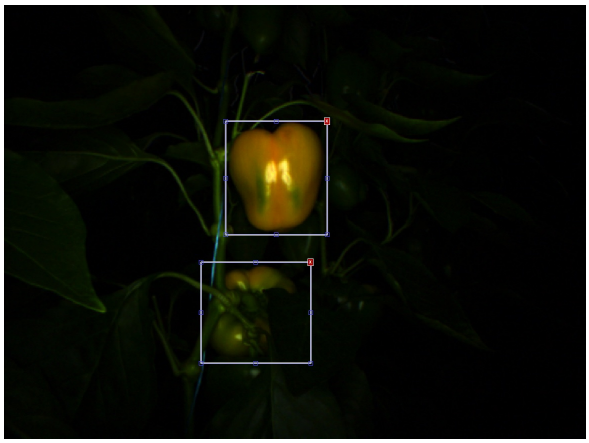


Figure 4.9: Labeling process - window based manual tagging.

Figure 4.10: Same image taken with artificial illumination (left) and natural overexposed light (right).

The annotator was requested to tag the image by reviewing both the image with natural light (available in DB#1 and DB#2) and with artificial illumination (Figure 4.10), resulting in a single merged ROI.
The mutual number of targets detected from multiple viewpoints was also evaluated by manual review of each couple of images taken from two viewpoints, and manual assessment of the number of joint targets seen from each combination. The output of the tagging is a set of viewpoint combinations and the number of joint targets detected.

4.4.3 Minimal detectability of annotators

Bounding box areas A_l and relative bounding box area RA_l (Equation 4.4) were calculated for each annotation of DB#2 and DB#3 (Figure 4.11). The minimal RA_l for DB#2 was 0.7% which is equivalent to a bounding box area A_l of 13440 px. The minimal RA_l for DB#3 was 0.21% which is the equivalent to a bounding box area A_l of 1024px.

4.4.4 Single viewpoint results

The average detectability of each viewpoint across all scenes acquired is presented in Figure 4.12. Early in the harvesting season a single viewpoint provides between 40-90% detectability. In mid-season 40%-60% of the fruit are detected from one viewpoint. At the end of the season only 20-30% fruit are detected. However, the average detectability of the different viewpoints

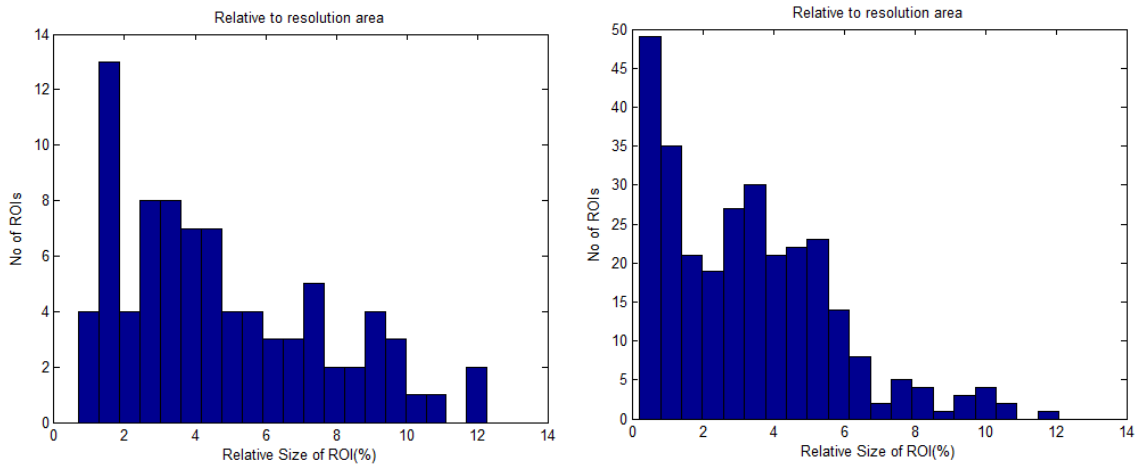


Figure 4.11: Distribution of RA_l in DB#2 (Left) and DB#3 (Right).

is 60%-70%, both at the beginning of the season and in the middle of the season, indicating the importance of finding the best viewpoint.

Detectability results for DB#2 are significantly lower than for DB#1 and DB#3, as expected, due to the harsh environmental conditions caused by the fact that it was later in the season, where the occlusion is maximal. Additionally, the acquisition was shortly after the previous harvesting and the sensing rig was randomly placed along the lane, resulting in a significantly lower number of detected targets, as well as lower detectability rates.

Detectability analysis as a function of the viewpoint features: distance (Table 4.3), tilt (Table 4.4) and azimuth angles (Table 4.5), are presented. Results show higher detectability with an increase in the distance to the target, which can be explained by a wider overview taken at a greater distance. The inconsistent rise of detectability as a function of distance in DB#3 can be explained by the uneven number of viewpoints taken from each distance according to the protocol (i.e., one viewpoint from 17cm as opposed to four viewpoints from 19cm). Positive tilt, in which the camera faces upwards, resulted in better detection results than negative tilt, and in some cases in results equivalent to detection results at 0 tilt. This supports previous findings [Hemming et al., 2014b] that suggested that optimal viewpoints are those facing upwards. Variations in the azimuth angle showed an inconsistent effect on detectability. In DB#1 no azimuth was preferred, while in DB#2 both positive and negative azimuth angles equally increased detection. Finally, in DB#3 positive azimuth angles gathered significantly better results than negative azimuth angles.

A closer look into the correlation between the lane side and the azimuth angle, described for DB#2 (Table 4.6), reveals that the preference to positive or negative azimuth dependents on the side of the lane. For the side of the lane with peppers facing west viewpoints with positive azimuth angles have a better detection rate than viewpoints with negative azimuth angles. For the side of the lane where the peppers are facing east, the detection rate for negative azimuth angles is lower than for positive azimuth angles. These results are also supported by DB#3, which yields better results for positive azimuth angles. The negative azimuth angle value seems to be too large for the given conditions, providing significantly lower detection results. DB#3 yielded similar detection results for both sides of the lane,

Table 4.3: Detectability results, single viewpoint by distance.

DB#1		DB#2		DB#3	
Distance (mm)	Detection (P_i)	Distance (mm)	Detection (P_i)	Distance (mm)	Detection (P_i)
300	52%	400	19%	70	61%
350	58%	700	27%	80	56%
450	61%			90	64%
				100	55%
				170	76%
				190	74%

Table 4.4: Detectability results, single viewpoint by tilt angle.

DB#1		DB#2		DB#3	
Tilt (deg)	Detection (P_i)	Tilt (deg)	Detection (P_i)	Tilt (deg)	Detection (P_i)
-22	56%	-30	5%	-20	46%
0	54%	0	29%	0	63%
22	64%	30	23%	10	56%
				20	79%

which can be explained by the presence of the artificial growing lights that evened out the natural orientation of the lane, as opposed to DB#2. Unfortunately, data on lane side of DB#1 were not part of the acquisition protocol.

4.4.5 Weighted viewpoint score results

Results of the weighted viewpoint score, Equation 4.4, as function of the weights and distance are presented in Table 4.7. Results show expected lower scores for vectors with lower values of $W_F + W_T + W_O$, but consistency in rise in score with higher distance, found for detectability P_i from the i's viewpoint. Results of the weighted score as function of tilt angle (Table 4.8) prioritize positive tilt to negative tilt, for all weights. Results of the weighted score as function of azimuth angle (Table 4.9) show preference to positive azimuth angles consistently for all weight vectors.

4.4.6 Combined viewpoints results

The results of the combined viewpoint analysis performed on DB#1 (Table 4.10) suggest that the right combination of viewpoints can raise the detection from the range of 40%-60% for a single viewpoint up to 85% for a combination of viewpoints. The best results are achieved by combining different sets of tilt and azimuth angles. For example, the combination of viewpoints 2 and 4 (Table 4.2) yields the best results in this case (85%); the combination of

Table 4.5: Detectability results, single viewpoint by azimuth angle.

DB#1		DB#2		DB#3	
Azimuth (deg)	Detection (P_i)	Azimuth (deg)	Detection (P_i)	Azimuth (deg)	Detection (P_i)
-45	47%	-50	27%	-50	54%
0	62%	0	20%	0	64%
45	51%	50	27%	20	75%

Table 4.6: Detectability results, single viewpoint by azimuth angle and lane side for DB#2, DB#3. For DB#2 lane sides (1) West (2) East. For DB#3 lane side (1)Southeast (2) Northwest.

	DB#2				DB#3			
Azimuth/ Lanes side	-50	0	50	Total	-50	0	20	Total
1	31%	26%	19%	26%	52%	66%	67%	65%
2	21%	12%	38%	19%	56%	63%	80%	63%

Table 4.7: Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and distance for DB#3.

	Weighted viewpoints score(S_i)	
Distance (mm)	$W_F = 1\ W_T = 1\ W_O = 0$	$W_F = 1\ W_T = 0.7\ W_O = 0.5$
70	61%	42%
80	56%	41%
90	64%	40%
100	55%	39%
170	76%	57%
190	74%	54%

the distance of 450mm from different tilt angles ($0°$ and $22°$) and different azimuth angles ($45°$ and $0°$) yields the best results. However, it must be noted that none of these viewpoints are the best single viewpoint, which is viewpoint 1 (69% detectability). The detection P_i of those viewpoints, when analyzed as a single viewpoint (59% and 65%, accordingly, for viewpoints 2 and 4), do not yield the best detectability results. However, their combination yields the best results, probably due to the variation in the tilt and azimuth angles. On the other hand, the combination of viewpoint 7 and viewpoint 12 yields the worst results in this case (53%) and are, according to the protocol, a combination of the same tilt angle ($0°$) and azimuth angles ($-45°$) but a variation in distance (300mm and 350mm, accordingly). In this case, the simple change in distance with no change in angles yields very little additional detectability, since viewpoint 7 alone provides 43% detectability while viewpoint 12 alone provides 46% detectability.

Analyzing the fruit detectability dependence on distance to fruit suggests that combinations of higher distances provide better detection results, while repeating the same higher distance is better than a combination of lower distances (Table 4.11). Analysis by tilt angle (Table 4.11) suggests that the combination of positive tilt and 0 tilt yields best detection results. Analysis by azimuth angle (Table 4.11) suggests that the combination of no azimuth with a positive azimuth provides the best detection.

Table 4.8: Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and tilt angle for DB#3.

	Weighted viewpoints score(S_i)	
Tilt (deg)	$W_F = 1\ W_T = 1\ W_O = 0$	$W_F = 1\ W_T = 0.7\ W_O = 0.5$
-20	46%	36%
0	63%	44%
10	56%	43%
20	79%	55%

Table 4.9: Weighted viewpoint score as function of weights (W_F – fully visible weight, W_T – truncated weight, W_O - occluded weight) and azimuth angle for DB#3.

	Weighted viewpoints score(S_i)	
Azimuth (deg)	$W_F = 1\ W_T = 1\ W_O = 0$	$W_F = 1\ W_T = 0.7\ W_O = 0.5$
-50	54%	41%
0	64%	46%
20	75%	51%

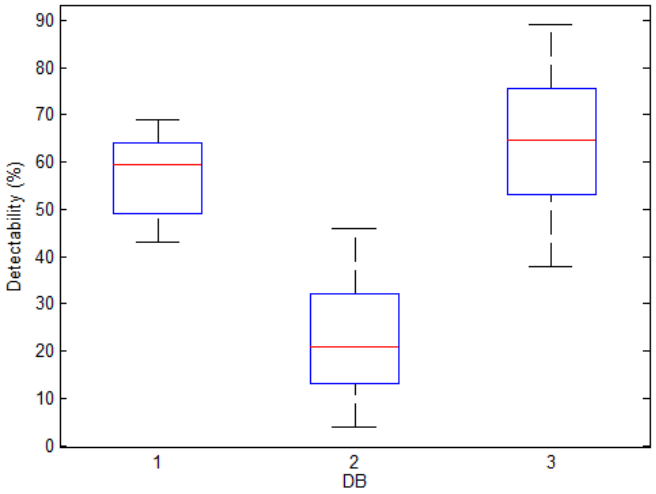


Figure 4.12: Detectability results, single viewpoint.

These results are similar to the single viewpoint results. Greater distance and positive tilt yield better detectability in both single and combined viewpoints. The combined viewpoints analysis provides additional insights, suggesting combining viewpoints with different angles for a better overview of the scene. Combining different distances does not provide significant additional detectability, and higher distance of any viewpoint is always preferable. This is though limited both by the greenhouse size and the resolution of the camera. Therefore, this conclusion is defined under the assumption that the camera controllable parameters remain static and that the variation of distances is within the close range to the distances presented, to which these conclusions can be extrapolated.

4.4.7 Detectability modeling results

The results of the detectability model for DB#1 (Table 4.12) support the descriptive results, with statistically significant correlation between distance and number of detected peppers (p.value 0.045), as well as near statistically significant (when using a significance level of 5%) correlation for tilt (p.value 0.061). Both variables present positive correlation with the number of detected peppers, implying that for higher distance and tilt, the expected number of detected peppers is higher. The results of DB#2 (Table 4.12) show significance for lane side with positive correlation (implying a higher number of peppers detected when the peppers are facing east) and for the interaction between the distance and the tilt. While this interaction was found to be statistically significant, the regression weight is negligible. For DB#3 (Table 4.12), all three controllable variables revealed a statistically significant positive relation.

Table 4.10: Detectability results, combined viewpoints for DB#1.

VP	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	69%	84%	77%	78%	78%	79%	78%	79%	77%	78%	82%	76%	81%	80%
2		59%	73%	85%	77%	68%	74%	80%	78%	79%	67%	76%	82%	84%
3			50%	70%	71%	69%	60%	70%	68%	74%	70%	60%	71%	73%
4				65%	77%	77%	75%	74%	74%	79%	78%	73%	73%	79%
5					61%	70%	70%	75%	65%	67%	73%	69%	75%	70%
6						46%	64%	70%	64%	74%	54%	64%	76%	72%
7							43%	73%	65%	74%	68%	53%	73%	73%
8								60%	68%	76%	73%	71%	70%	73%
9									51%	67%	66%	66%	73%	67%
10										65%	77%	72%	78%	71%
11											49%	68%	77%	75%
12												46%	71%	71%
13													66%	76%
14														64%

Table 4.11: Detectability results, combined viewpoints for DB#1 by distance, tilt and azimuth.

Distance (mm)	300	350	450
300	65%	69%	74%
350		70%	76%
450			74%

Tilt (deg)	-22	0	22
-22	62%	71%	74%
0		69%	76%
22			69%

Azimuth (deg)	-45	0	45
-45	54%	72%	70%
0		73%	76%
45			59%

In order to retain the discovered relation between the distance and tilt, and the number of detected peppers, a minimum of 12-16 scenes (40%-50% training ratio) is needed for DB#1 (Table 4.13). A smaller dataset, results in a loss of the ability to determine these features as significant. The minimum number of scenes required for DB#3 is more challenging to define since it showed an inconsistent rise in p.values for smaller datasets. This could be related to the overall small number of scenes in DB#3 (8) compared to DB#1 (32). Nevertheless, a minimum of 4 scenes (50% training ratio) is the critical number of scenes for which none of the features initially found significant in the overall dataset are significant any longer. DB#2 was not reviewed in this setting, due to the low significant results presented in the detectability modeling on the whole dataset presented in Table 4.12.

The goodness of fit feature RMSE (Table 4.13), which aims to evaluate the ability of the model to predict the number of detected peppers, revealed an expected growth in error for smaller training datasets for DB#1 and DB#3. The range of values for the RMSE of DB#3 is higher than for DB#1. This can be explained by the smaller dataset in DB#3 and the higher variance of the overall number of targets visible in each scene in DB#3 (std=1.504) compared to DB#1(std=0.932). The values of the RMSE for all training-testing sets of DB#1 are significantly lower than the standard deviation, implying that the prediction model is significantly better than a $\bar{T}$ model which constantly predicts only the average number

Table 4.12: Modeling results, including significant variables and interactions, and regression weights.

		DB#1		DB#2		DB#3	
		Regress. weight	p.val	Regress. weight	p.val	Regress. weight	p.val
Control. var.	Distance	0.001	0.045**	-	>0.1	2.38	0.042**
	Tilt	0.006	0.061*	-	>0.1	0.009	0.022**
	Azimuth	-	>0.1	-	>0.1	0.005	0.043**
	Lane side	NA	NA	0.668^a	0.015**	NA	NA
Inter. var.	Distance*Tilt	-	>0.1	$<10^{-4}$	0.039**	-	>0.1
	Distance*Azimuth	-	>0.1	-	>0.1	-	>0.1
	Distance*Lane side	NA	NA	-	>0.1	NA	NA
	Tilt*Azimuth	-	>0.1	-	>0.1	-	>0.1
	Tilt* Lane side	NA	NA	-	>0.1	NA	NA
	Azimuth* Lane side	NA	NA	-	>0.1	NA	NA

[a] - Peppers facing east; *p.val<0.1; **p.val<0.05

Table 4.13: Prediction model accuracy and significant variables at random train-test split.

	DB#1					DB#3					
Train /test ratio	No. train scenes	Dist p.val	Tilt p.val	RMSE	IOA	No. train scenes	Dist. p.val	Tilt p.val	Azim. p.val	RMSE	IOA
90/10	28	0.099*	0.038**	0.849	0.342	NA					
80/20	25	0.054*	0.084*	0.749	0.358	6	0.041**	0.097*	0.235	1.538	0.57
70/30	22	0.14	0.266	0.764	0.2	5	0.193	0.026**	0.086*	1.646	0.477
60/40	19	0.199	0.026**	0.957	0.236	NA					
50/50	16	0.192	0.086*	0.861	0.273	4	0.427	0.067*	0.101	1.632	0.515
40/60	12	0.026**	0.050**	0.918	0.339	3	0.108	0.325	0.177	1.462	0.504
30/70	9	0.447	0.778	0.998	0.251	2	0.522	0.19	0.040**	1.495	0.564
20/80	6	0.609	0.134	0.935	0.268	1	0.189	0.969	0.313	1.656	0.525
10/90	3	0.307	0.941	1.005	0.362	NA					

*p.val<0.1, **p.val<0.05

of peppers. This suggests that the defined variables do indeed make the prediction more accurate. DB#3 did not yield similar results, suggesting that the size of the dataset is too low for an accurate perdition model.

The IOA values remained relatively static for all train-test splitting ratios, presenting significantly higher values for DB#3 than for DB#1. This suggests that the prediction ability with the variability presented in DB#3 is significantly higher than in DB#1 due to the complexity of the scene in the middle of the growing season compared to complexity at the beginning of the season.

Temporal analysis performed on DB#3 (Table 4.14) resulted in slightly lower RMSE values compared to the random split dataset, and similar IOA and spatial variable significance values. This implies that prediction in a temporal manner is slightly better, but the training sets were too small to indicate if these results are significant.

4.5 Conclusions

This research provides a methodology to model fruit detectability based on data acquired automatically in the field. The methodology includes a protocol for automatic data acquisition

Table 4.14: Prediction model accuracy and significant variables at temporal train-test split, DB#1. Average values across all possible combinations.

No of train scenes	No of test scenes	Distance p.value	Tilt p.val	Azim. p.val	RMSE	IOA
7	1	0.073	0.019	0.066	1.498	0.489
6	1	0.186	0.011	0.1	1.494	0.51
5	1	0.321	0.035	0.109	1.292	0.543
4	1	0.404	0.068	0.174	1.328	0.555
6	2	0.167	0.01	0.119	1.393	0.529
5	2	0.301	0.019	0.115	1.301	0.57
4	2	0.426	0.074	0.199	1.317	0.556

to ensure data includes the object, environment and robot variability. Although the developed model was applied only for sweet pepper data analyses, it can be applied to model detectability of other fruit varieties.

4.5.1 Viewpoints

Choosing the right viewpoints or combination of viewpoints for detection is crucial to provide high detectability rates in autonomous harvesting systems. For sweet pepper harvesting, at the beginning of the season it is crucial to determine the right viewpoint a-priori based on analyses on real data in real world conditions, since the difference between the detectability from the best viewpoint and the worst is up to 50%. In these less complex environmental conditions, a single best viewpoint can provide detection of up to 90%; therefore, it might be sufficient to perform detection with only one viewpoint. In mid-season the variation between viewpoints is only 20%. This means that it is still crucial to find the best viewpoint, but even the best one will provide only up to 70% detectability, and the key for sufficient detectability for robotic harvesting lies in combining multiple viewpoints. Hence, it is important to employ the methods presented in this chapter to determine the best combination of viewpoints.

Distance from fruit has significant influence on detectability, with greater distance resulting in better detectability. However, this of course depends on sensor and robot specifications (e.g., resolution, reachability) and on greenhouse conditions (e.g., lane width, plant density). Moreover, the distance cannot be used as a sole predictor of detectability, since the same distance revealed different detectability rates in different growing conditions.

Positive tilt angles yields better detectability. In some of the environmental conditions, the differences between no tilt and positive tilt were not very significant. The azimuth angle does not influence detectability as an independent parameter. However, when analyzed in combination with the lane side, better results are achieved with positive azimuth angles for peppers facing east and with negative azimuth angles for peppers facing west. This can be explained by the direction of leaf growth in relation to the sun.

Weighted viewpoints score results resulted in similar viewpoints for improved detectability. Therefore, the recommended viewpoints in the case study of sweet peppers remain those with positive tilts, higher distance and in some cases positive azimuth.

The combined viewpoint analysis revealed that in mid-season 85% of the fruits can be detected for a combination of viewpoints; this represents a significant increase in detectability from 40%-60% from a single viewpoint. In-depth analysis suggests that a combination of viewpoints

with different spatial angles (tilt and azimuth) provides better detectability. The highest chosen distance should remain static, since a combination of lower and higher distances does not provide additional detectability compared to the higher distance alone.

4.5.2 Detectability modeling and training set minimal size

The statistical modeling results support the conclusions of the descriptive statistics, where, in most cases, distance and tilt yielded a positive relation to detectability. The data from DB#2 showed too harsh conditions to enable proper modeling, with too low detection rates to statistically model detectability.

The prediction results of the detectability modeling revealed that a minimum of 40%-50% training ratio (~ 12 scenes) is necessary in order to be able to discover the statistically significant spatial variables. Beginning of the season measurements in DB#3 showed better predictability results, due to the less occluded scenes present at the beginning of the season and the consistent environmental variables, such as orientation of the lane and placement in the greenhouse.

Temporal analysis of DB#3 revealed a slight reduction in RMSE, but similar results to the RMSE of random split. This suggests that the size of the dataset was too low for significant environmental differences between the first and the last scenes.

4.5.3 Applicability of the results

This research provides an important step towards improving detection rates, which are currently a bottleneck to robotic harvesting, by finding the best viewpoints to increase detectability. Results indicate in which conditions multiple viewpoints are necessary in order to increase detection rates. The specific viewpoints for sweet pepper harvesting were determined in a systematic manner.

The ground truth measured in this research is based on a human annotator that marks acquired images, but eliminates a need of a human in the field for ground truth acquisition which is a tedious task. With the advancement of learning algorithms [McCool et al., 2016; Sa et al., 2016] ideal detection can become feasible and enable to replace human annotators for providing ground truth.

This research also provides a methodology for identifying controllable variables that significantly affect the detectability rate. Using the defined protocols similar detectability models can be applied to other robotic harvesters aiming to provide prediction of detectability given a set of controllable variables that are manageable by the provided hardware. If a given harvester is equipped with additional degrees of freedom such as dynamic lens or artificial lighting these can also be included into the model by following the protocol described.

In this work natural illumination was not noted as a measurable variable. The values were subject to extreme changes and therefore manual measurements were not feasible. In a system equipped with an autonomous natural light sensor the influence of it should be evaluated along with development of algorithms that are robust to natural light fluctuations such as

flash cut techniques [Sun et al., 2007]. Since in most research to date controlled illumination was used to overcome the natural illumination variability this was defined as a controllable variable.
The developed methodologies can be applied to improve robotic harvesting of other crops in dense environments. Ongoing work aims to include the findings regarding the best viewpoints to improve the harvesting performance of a sweet pepper harvesting robot as a continuation of the European FP7 project Clever Robots for Crops - CROPS[1] [Bontsema et al., 2015; Hemming et al., 2014a], and within the current European Horizon 2020 project SWEEPER [Ringdahl et al., 2017].

4.6 Research questions answered

The following research questions defined in Section 1.2 have been met:

RQ1.1: What are the static best characteristics of a sensing viewpoints for sweet pepper harvesting? Are there any at all? Guidelines for preferred viewpoints features have been outlined for the given use case of sweet pepper harvesting using the outlined robotic configuration. Nevertheless, no global best viewpoint has been found and therefore an algorithm for dynamic viewpoint selection has been designed and is described in Chapter 5.

RQ1.2: How do the preferred viewpoint characteristics change along the season, growing conditions, cultivation techniques, and varieties? Common characteristics of better viewpoints have been defined along seasons (e.g. tilt angle). However, the conclusions differ between growing conditions e.g., beginning of season detectablity is high enough for a single viewpoint while mid to end of season require combination of viewpoints.

RQ1.3: What methodologies should be applied to perform a similar analysis for a different crop variety/conditions? Methods have been outlined along with an accommodating flow chart (Figure 4.1).

RQ1.4: What is the minimal dataset needed for dynamic learning of best viewpoint characteristics for a given setting? Minimal dataset sizes were derived for specific statistically significant features through power analysis.

RQ1.5: Are there temporal relations between subsequent plants along the row in terms of detectability and best viewpoint characteristics? Some temporal relations were found but the results were insignificant in the given conditions.

[1] http://www.crops-robots.eu/

The use of dynamic sensing strategies to improve detection

5

> *There are things I can't force. I must adjust. There are times when the greatest change needed is a change of my viewpoint.*
>
> **— Denis Diderot**
> (French philosopher, art critic, and writer.)

- Published in Kurtser, Polina & Yael Edan (2018). "The use of dynamic sensing strategies to improve detection for a pepper harvesting robot". IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).
- **Research objective RO2:** Development of a dynamic sensing algorithm that will predict the need of an additional viewpoint and its location.

This chapter presents the use of dynamic sensing strategies to improve detection results for a pepper harvesting robot. The algorithm decides if an additional viewpoint is needed and selects the best-fit viewpoint location from a pre-defined set of locations based on the predicted profitability of such an action. The suggestion of a possible additional viewpoint is based on image analysis for fruit and occlusion level detection, prediction of the expected number of additional targets sensed from that viewpoint, and final decision if choosing the additional viewpoint is beneficial. The developed heuristic was applied on 96 greenhouse images of 30 sweet peppers and resulted in up to 19% improved detection rates compared to conventional single viewpoint sensing. The harvesting utility cost function decreased by up to 10% compared to the single viewpoint strategy.

5.1 Introduction

Commercial applicability of autonomous harvesting robots indicates a need to detect at least 90% of the actual number of fruits in the field [Blackmore et al., 2002]. To date most work in agricultural robotics relied on detection from a single viewpoint. However, the number of actual fruits visible from a single viewpoint (defined as the visibility from a single viewpoint) in the agricultural environment is known to be limited to 40%-60% [Kurtser & Edan, 2018a; Hemming et al., 2014b; Bulanon et al., 2009]. The term of visibility from a single viewpoint should not be confused with the standard performance measures of image detectability applied in the computer vision community, which defines true positive and true negative rates by comparing the number of objects detected by the algorithm to the number of objects in the image as labeled by a human annotator [Russell et al., 2008]. Current algorithms have limited 85-90% true positive rates (also noted as detection rates, [Vitzrabin & Edan, 2016a; Bac et al., 2014b]) identified as a major bottleneck limiting agricultural robots' performance [Bac et al., 2014b].

The low visibility rates are mostly due to the dense agricultural environment (e.g., occlusions by leaves [Ringdahl et al., 2018], Figures 5.1 and 5.2). Attempts to cope with occlusion have been made using hyperspectral imaging [Lass & Prather, 2004], but the high cost of hyperspectral cameras as well as their weight has caused RGB cameras to be the most widely used cameras in robotic harvesting [Bac et al., 2014b]. Additional solutions to overcome occlusions include mechanical removal in a temporal manner using air blowing [Edan et al., 2000; Yoshida et al., 1985], and permanent pruning of leaves. Both methods can cause reduced yield and damage the plant, and hence are preferably avoided.

In this chapter we investigate an alternative way to improve detectability for high-density crops with low visibility by combining multiple viewpoints [Kurtser & Edan, 2018a; Hemming et al., 2014b; Bulanon et al., 2009] to increase visibility and thereby detection rates.

The search for optimal viewpoints has been extensively investigated in several domains (e.g., computational geometry, visual servoing, robot motion, graph drawing, grasping, medical intervention planning, and multi-camera stereo vision [Krainin et al., 2010; Mühler et al., 2007; Vázquez et al., 2001]). These methods often rely on generating image-based models from scenes with known [Reed & Allen, 2000] or unknown [Bajcsy, 1988] geometry. Often, the underlying assumption is the existence of an analytical model to support calculation of the best viewpoint. The viewpoints are calculated to obtain optimal information given some initial model. As a result the selected viewpoints are often static and predefined.

Previous research that analyzed viewpoints for sweet pepper detectability [Kurtser & Edan, 2018a] revealed that due to the inherent biological variability no specific viewpoint provides best detection in all cases. However, multiple viewpoints provide increased detectability by providing more information. Therefore, in this chapter we present a new approach for agricultural robots that uses a dynamic sensing strategy.

Dynamic selection of viewpoints is regarded as active sensing, active perception, dynamic sensing, or next best viewpoint selection algorithms. Classical research in the active sensing field [Bajcsy, 1988] assumes that "in scenarios where data measurements are relatively expensive or slow, we want to know where to look next so as to learn as much as possible..." [MacKay, 1992], where information is often measured as viewpoint entropy [Vázquez et al., 2002]. This method assumes all additional information is valuable and does not take into account the cost of the additional sensing versus the benefit from it. The information gathered by an additional viewpoint, suggested by a maximum information policy, might bring too little profit compared to the cost of the sensing operation and therefore undermine the overall performance. The sensing operation's costs are caused by the time of the additional path travelled by the robotic manipulator to and from the viewpoint and the actual additional sensing time, which both increase cycle times. The extra sensing is expected to provide improved detection by overcoming occlusions from an additional viewpoint.

In this research we present an algorithm that decides if an additional viewpoint is needed and heuristically selects the best-fit viewpoint location from a pre-defined set of locations based on the predicted profitability of such an action. Similar algorithms often rely on viewpoint entropy [Vázquez et al., 2002] as a measure for information content; however, this might be problematic in the unstructured dynamic agricultural environment, where

scene modeling is very limited [Kurtser & Edan, 2018a]. Therefore, we introduced a more domain-related heuristic measure of information content. Since additional sensing increases the cycle time, another limiting factor of harvesting feasibility [Elkoby, 2016; Elkoby et al., 2014], an additional viewpoint should be pursued only if the additional sensing provides valuable enough information. Hence a possible additional viewpoint is proposed by predicting the expected number of additional targets sensed from that viewpoint based on image analysis for fruit and occlusion level detection, and deciding if the additional viewpoint is beneficial, taking into account the value of these additional targets versus the cost of their acquisition. The viewpoint is selected from a pre-defined set of locations to simplify the robotic operations and limit reachability issues in the dense agricultural environment where motions are limited.

The objective of this chapter is to **propose a dynamic sensing algorithm** that **decides if an additional viewpoint is profitable** based on the information derived from the current viewpoint, and **if so, directs the robot to the best-fit viewpoint** from a set of predefined list of possible viewpoints. To the best of our knowledge, active/dynamic sensing has been limitedly investigated in the agricultural domain [Foix et al., 2015] and has not been applied to robotic harvesting. The proposed algorithm is evaluated for a case study of sweet pepper harvesting.

5.2 Dynamic sensing algorithm

The algorithm includes four main steps: (i) analyze the image to detect fruit and occlusion level; (ii) suggest a possible additional viewpoint; (iii) predict the expected number of additional targets sensed from that viewpoint; (iv) decide if choosing the additional viewpoint is beneficial.

5.2.1 Image analysis of the scene

The targets (fruits) and occlusions (leaves and stems) detection algorithm includes three empirically developed steps to derive the number of fruits and calculate the occlusion level in the current scene.

Image segmentation

A k-means clustering algorithm [Lloyd, 1982] is applied on the transformed image (Figure 5.3) according to Equation 5.1.

$$\begin{aligned} Channel1 &= G - R \\ Channel2 &= B - R \end{aligned} \tag{5.1}$$

where R, G, and B are the RGB channels of the original image. K-means clustering is a common data-partitioning algorithm that assigns each observation to a predefined number of clusters based on the least distance between each observation and the centroid of each cluster. The cluster containing a centroid with the lowest values of the $B-R$ channel is defined as the peppers cluster, the cluster with a centroid with the highest values of $B-R$ is the background cluster, while the cluster in between corresponds to the occlusion cluster.

Figure 5.1: Examples of visibility variability in different crop types. Left: Cherry tomatoes, high visibility. Right: Grapes, mid/low visibility.

Figure 5.2: Image taken in a greenhouse including 3 peppers, frontal leaves, side leaves, and additional occluding items.

As a result, the image (Figure 5.2) is segmented into three clusters - fruits, occlusions, and background (Figure 5.4).

Number of fruits

The classified peppers cluster goes through a filtration procedure of all connected elements by thresholding based on their area, aspect-ratio, and the average intensity of the hue level of the element in the original image (e.g., Figure 5.5). The number of unconnected elements remaining in the cluster after this process is defined as the number of revealed peppers, N_{RP}.

Occlusion level calculation

A two-step filtering process is performed to separate frontal leaves and side leaves and stems (Figure 5.7). First, the image is cleared of noise (such as the pepper edges and small artifacts) using erosion-dilation-hole filling procedures (Figure 5.6). Then, the cluster is filtered based on two values calculated for each unconnected element in the mask: the area and the aspect-ratio.

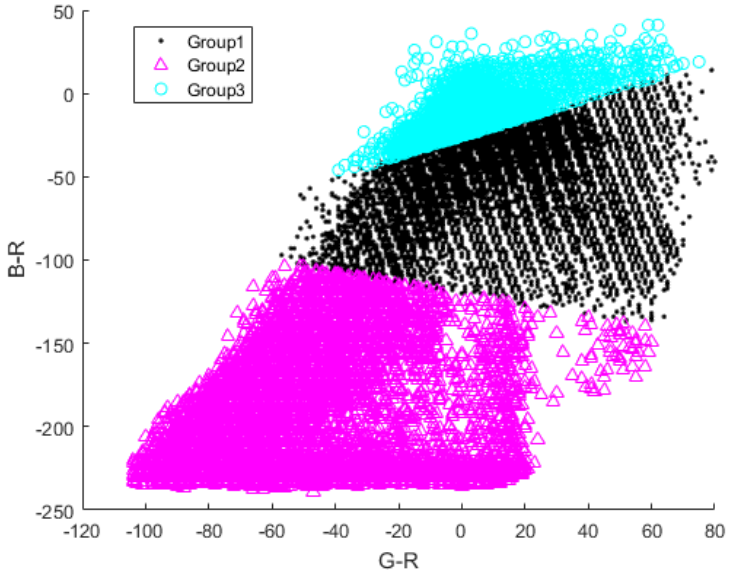


Figure 5.3: Results of k-means classification on a transformed RGB image. Groups 1-3 represent the classification into peppers, occlusion and background.

The occlusion level (OL) of the image (Equation 5.2) is defined as the ratio between the occluded area and the overall area in the scene.

$$OL = \frac{\sum_{i=1}^{N_0} A_i}{A_S} \tag{5.2}$$

where A_i is the area of the occluding element i, A_S is the area of the obtained image, and N_o is the number of occluding elements. The value of OL is estimated using the obstacle mask generated: A_i is calculated by summing the produced binary mask of occlusion, N_o is the number of unconnected items within the mask, and A_S is static for a given camera resolution.

5.2.2 Suggestion of an additional viewpoint

Since previous research [Kurtser & Edan, 2018a; Hemming et al., 2014b] revealed that no viewpoint is consistently best in greenhouse conditions and that two consecutive scenes have very little resemblance in terms of best viewpoint, a heuristic method was developed to find the next viewpoint. The method is based on general guidelines for improving detection

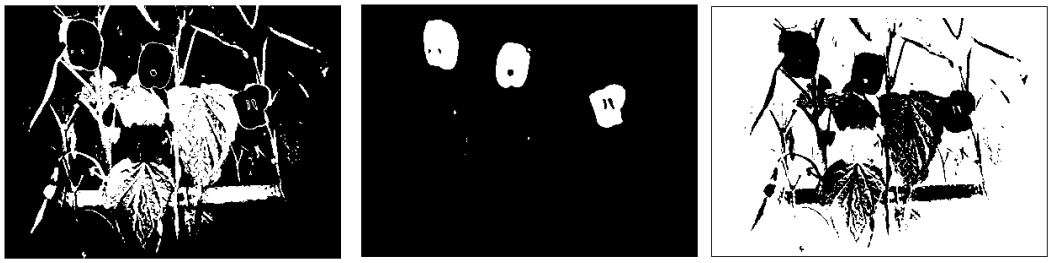


Figure 5.4: Three clusters classified by the k-means procedure. Left: occlusion, middle: peppers, right: background.

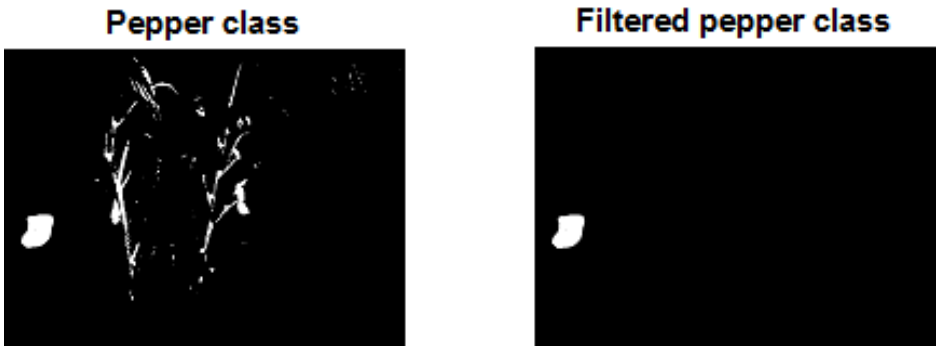


Figure 5.5: Resulted filtered pepper mask. $N_{RP} = 1$.

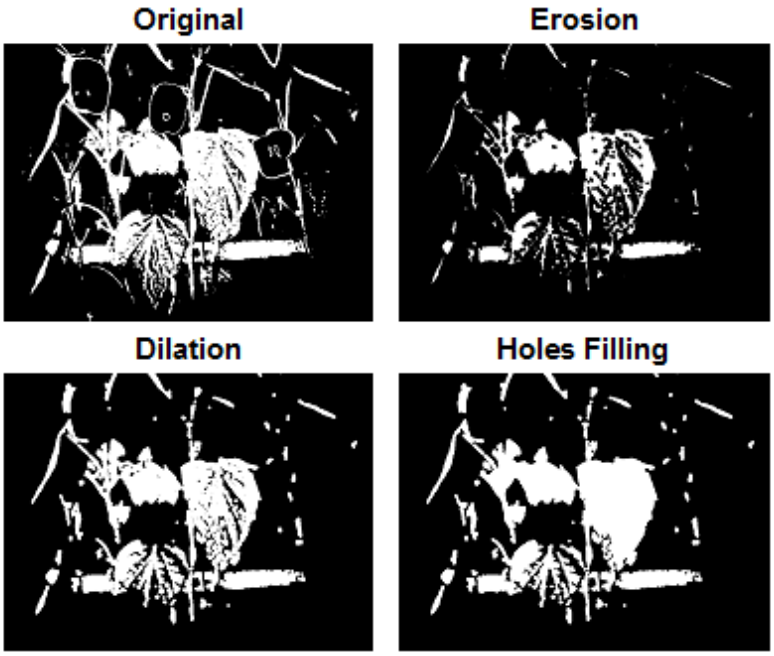


Figure 5.6: Noise reduction procedure on occlusion cluster.

Figure 5.7: Threshold by area and aspect-ratio.

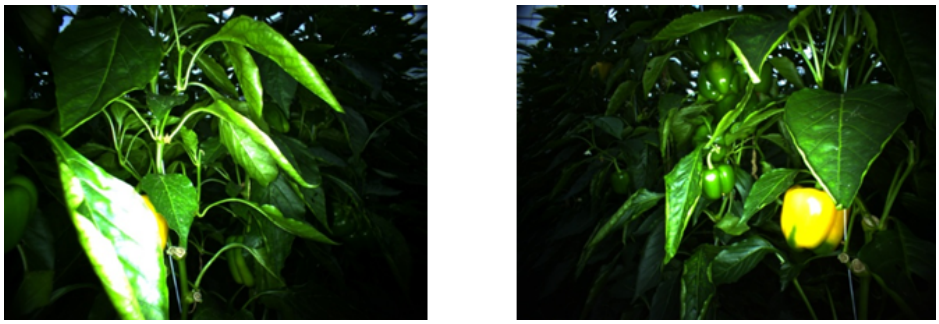


Figure 5.8: Change in yaw angle in highly occluded scenes. Yaw angles left: $45°$, right: $135°$.

Figure 5.9: Same scene with 3 viewpoints decreasing in distance between camera and stems from left to right. Left: 19cm, middle: 10cm, right: 7cm.

revealed in previous research [Kurtser & Edan, 2018a]. Two causes for undetected peppers are dealt with: frontal occlusions and camera field-of-view limitations. An occluding leaf sensed from one angle will no longer be occluding from another angle (Figure 5.8). Missed peppers caused by a limited field of view might be detected by increasing the distance of the sensor (e.g., camera) to the stems (Figure 5.9). Previous research [Kurtser & Edan, 2018a] indicated that a greater distance of the sensor (assuming the same lens configuration is used) improves detection up to the limit of image resolution. But keeping a greater distance from the crop at all times causes additional fruit approach time from the overview point [Ringdahl et al., 2018; Ringdahl et al., 2017] causing an extended harvesting cycle, as well as reachability issues due to the configuration of the robotic arm, and potential harm of the lane behind the robot due to the tight space in which the robot operates. Hence, closer viewpoints were noted as preferable.

The additional viewpoint location is calculated based on the occlusion level revealed from the first viewpoint and the distance from the viewpoint to the target.

The decision process is as follows (Figure 5.10): if the image's occlusion level is low the robotic manipulator is moved away from the plant, retaining the angle but increasing the distance; for a highly occluded image, the yaw angle from which the scene is sensed is changed to provide fewer occluded scenes.

A predefined list of yaw angles (0, left, and right) and two distances (low and high) is used based on previous analyses in real world conditions [Kurtser & Edan, 2018a]. The value separating high and low occlusion levels is based on the median of the occlusion levels encountered in the dataset.

The performance of the suggested decision process is evaluated using the measures defined in Section 5.3.4.

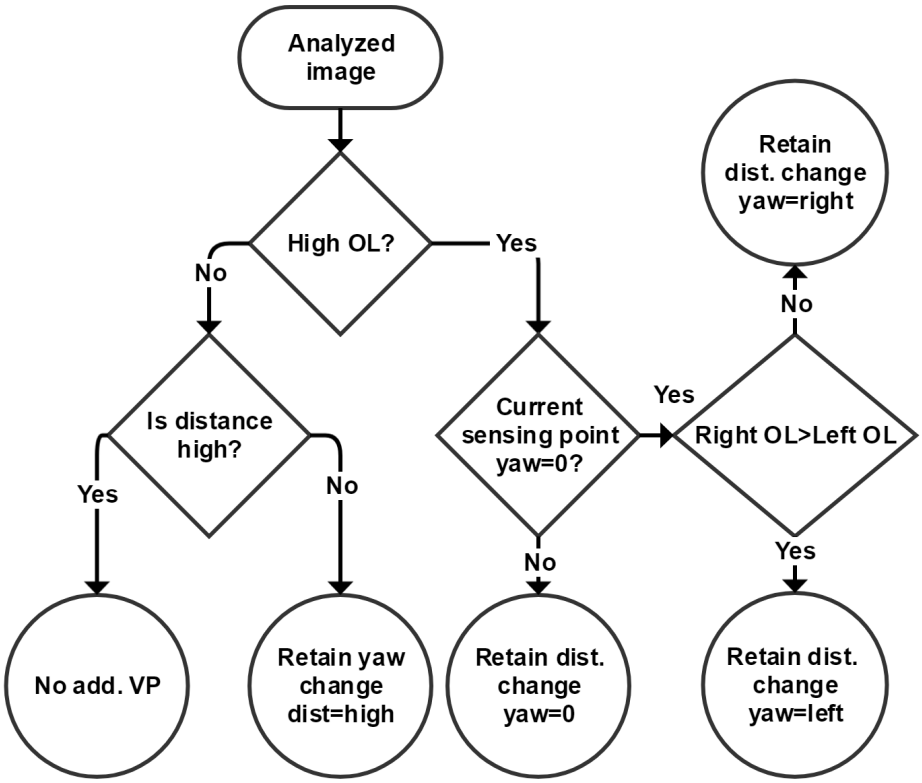


Figure 5.10: Next viewpoint decision tree.

5.2.3 Predicting the expected number of targets from the additional viewpoint

The number of targets to be revealed from a second viewpoint is predicted by deriving the number of sensed targets from the first viewpoint and modelling the expected scene yield from the second viewpoint.

The scene yield is assumed to be distributed as a spatial Poisson process [Privault, 2013] with an expected value of λ peppers within a scene. This is due to the nature of the problem where count data per area is used, and low numbers are present. The average scene yield λ depends on the growing conditions (e.g., the time of year, the variety of the pepper, distances between stems) and is derived based on ground truth information (Section 5.3.2).

The probability of a scene yielding N_P peppers is defined by the probability mass function of the Poisson distribution (Equation 5.3). The conditional probability of the scene to yield a total of N_P targets when N_{RP} targets has already been revealed is presented in Equation 5.4. The expected value of the scene yield is presented in Equation 5.5, and represents the expected number of overall targets in the scene given that N_{RP} targets have already been revealed.

$$P(N_P = k) = \frac{\lambda^k e^{-k}}{k!} \tag{5.3}$$

$$P(N_P = k | N_P \geq N_{RP}) = \begin{cases} \frac{\lambda^k e^{-k}}{k!(1-\sum_{i=0}^{N_{RP}-1} \frac{\lambda^i e^{-i}}{i!})} & N_{RP} \geq 1 \\ \frac{\lambda^k e^{-k}}{k!} & N_{RP} = 0 \end{cases} \tag{5.4}$$

$$E(N_P|N_P \geq N_{RP}) = \begin{cases} \frac{\sum_{k=N_{RP}}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!}}{1-\sum_{i=0}^{N_{RP}-1} \frac{\lambda^i e^{-i}}{i!}} & N_{RP} \geq 1 \\ \lambda & N_{RP} = 0 \end{cases} \quad (5.5)$$

The expected number of additional targets to be revealed from another viewpoint given N_{RP} depends on the detection probability P_D. The detection probability P_D is estimated according to the procedure described in Section 5.3.2. The expected value of targets to be revealed is defined in Equation 5.6.

$$E(N_R|N_P \geq N_{RP}) = E((N_P - N_{RP})\widehat{P_D}|N_P \geq N_{RP}) = \widehat{P_D}[E(N_P|N_p \geq N_{RP}) - N_{RP}] \quad (5.6)$$

5.2.4 Additional viewpoint decision

Sensing from an additional viewpoint is performed when the costs of sensing are lower than the potential loss as a result of undetected targets, and vice versa, an additional viewpoint is not executed when the cost of sensing is higher than the cost of potential undetected targets.

The **cost of an undetected target (C_{UT})** is the economic loss caused by a non-harvested fruit target. The **cost of a sensing operation (C_S)** is the cost associated with the additional sensing operation: the sum of the cost of the travel time of the robotic manipulator to the viewpoint and the sensing cost (time).

A sensing operation is performed when the condition in Equation 5.7 is satisfied.

$$E(N_R|N_P \geq N_{RP})C_{UT} > C_S \quad (5.7)$$

To simplify calculations, the sensing cost is assumed to be of constant value and independent of the viewpoint location. This assumption can be applied to robotic systems where the viewpoints are close and processing times are negligible.

5.2.5 Estimation of costs

While the economic calculation behind obtaining C_{UT} and C_S is out of the scope of this research, the algorithm assumes the presence of such costs. The cost of 1 sec of robot operation can be extracted given the price of the robotic harvesting system and the overall greenhouse profit. As a result, given that an additional viewpoint requires more time, this time can be converted into a financial loss.

Given a price for each pepper harvested, the costs of not harvesting a pepper or the cost of a human operator working side by side with the robot complementing for the peppers not harvested by the robotic system can be derived.

Initial calculations of such costs has been performed [Elkoby, 2016; Elkoby et al., 2014] but are highly variable between greenhouses. For manual pepper harvesting, the average yield reported by manual harvesters ranges between 0.6 and $0.8kg/m^2$ per day (13-16 harvesting hours) with an average of 3-9 seconds per pepper [Elkoby et al., 2014]. Assuming this is the pace of harvesting for a robotic harvester to be profitable, the cost of an unharvested pepper

in time units is assumed to be in the range of 3-9sec.
The cost of sensing, in time units, is defined as the distance to be travelled to a viewpoint. This largely depends on the robotic configuration. The additional viewpoint time cost C_S can be estimated as 0.4-1 sec, since the viewpoints in this experiment were located at distances between 20 and 50cm from each other, with an average robotic manipulator speed of 0.5 m/s for safety reasons. The processing time was assumed to be negligible. Therefore, for the given setting the C_{UT}/C_S ratio was varied between 3 and 20.
A 6 sec per pepper harvest cycle time was assumed to be profitable for the autonomous harvester based on analysis by the fp7 CROPS project[1] [Pekkeriet, 2011 and the HORIZON 2020 SWEEPER project[2]. This can be translated to a C_{UT}/C_S ratio of 6:1. Since other robotic harvesters might require a different ratio, a whole range of ratios were evaluated.

5.3 Evaluation methods

The described algorithms were evaluated on a dataset of sweet peppers (*Capsicum annuum*) images acquired in a greenhouse from several viewpoints, as well as ground truth information about the actual number of targets in each scene. The number of fruits visible from each viewpoint and also the number of fruits visible from each combination of two viewpoints (joint targets) were manually counted in the images. The proposed dynamic sensing algorithm was compared to an algorithm that included only a single viewpoint. Analyses were conducted for different sensing and undetected target costs.

5.3.1 Data collection and annotation

The database was acquired in a research greenhouse at Sint-Katelijne-Waver, Belgium, using a 6 degree of freedom manipulator (Fanuc LR Mate 200iD), equipped with an iDS UI-5250RE RGB camera to automatically acquire images from 12 viewpoints with artificial illumination conditions.
The spatial locations of the 12 viewpoints were defined by three parameters: distance to the plant, tilt angle, and azimuth angle (Figure 5.11). The viewpoints were selected as derived from the literature [Kurtser & Edan, 2018a; Hemming et al., 2014b; Bulanon et al., 2009] to ensure that half a sphere around the stem will be covered in the joint field of view [Kurtser & Edan, 2018a] and based on robotic reachability limitations only, as examined in a pretest in laboratory conditions. The acquisition protocol included manual centering of the sensory system in front of a pepper or a cluster of peppers to ensure correct ground truth measurements that were measured by manually counting the number of peppers in the scene. This also ensured that images with peppers are acquired for the analyses. Since peppers grow along the main stem and are planted at constant distances, this assumption is realistic for the robotic harvester, which can advance along the row in constant steps [Hemming et al., 2014a] or until the main stem is detected (e.g., [Bac et al., 2014b]).
A dataset with a total of 96 RGB images of 30 peppers, divided into 8 scenes with 2-6 peppers in each, was registered. The number of **detected targets** N_{RP} in each viewpoint was

[1]http://www.crops-robot.eu
[2]http://www.sweeper-robot.eu

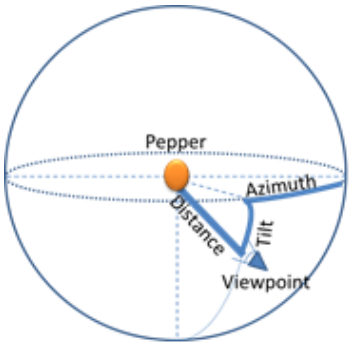


Figure 5.11: Definition of a viewpoint by distance and two angles – tilt and azimuth.

extracted either by manual annotation (denoted as N_{RP}^{M}), or based on the image analysis algorithm (denoted as N_{RP}^{A}).
The number of **joint targets** N_{JRP} was manually calculated for each pair of images by an annotator that either counted the joint number of targets from the original images (N_{JRP}^{M}) or counted the peppers in the automatically generated pepper masks of the original images (N_{JRP}^{A}). This enabled providing insights into the algorithm performance for both optimal detection rates provided by the human annotator (defined as the gold standard) and suboptimal detection algorithms provided by the automatically obtained image analysis number of peppers.

5.3.2 Parameters estimation

The reachable area is defined as the average number of reachable stems $\overline{S}$ multiplied by the average number of reachable clusters $\overline{C}$. This was derived based on field observations, which indicated that peppers usually grow on stems in two clusters, one above another. To estimate λ, the expected value of peppers in a scene Equation 5.8 is applied.

$$\hat{\lambda} = \overline{N_S}\,\overline{S}\,\overline{C} \tag{5.8}$$

where $\overline{N_S}$ is the average number of peppers on a stem and $\hat{\lambda}$ is the estimated value of λ. All parameters can be derived by either visually evaluating the scenes in a greenhouse prior to harvesting, or from botanical data gathered by the growers and from evaluation of images from previously collected scenes. The detection probability P_D is estimated by Equation 5.9.

$$\widehat{P_D} = \frac{\overline{N_{RP}}}{\hat{\lambda}} \tag{5.9}$$

where $\overline{N_{RP}}$ is the average number of detected peppers in all acquired images.

5.3.3 Sensing strategies evaluation and costs

Two sensing strategies are compared for each collected image from each viewpoint.

Single viewpoint

The total harvesting cost (TC) is calculated according to Equation 5.10.

$$TC = C_S + (N_P - N_{RP})C_{UT} \tag{5.10}$$

where N_P is the number of peppers in the scene as logged in the dataset and C_S and C_{UT} are the defined costs of sensing and undetected targets.

Dynamic sensing

If an additional viewpoint is not suggested, no additional costs are calculated and the total cost remains as described in Equation 5.10. If an additional viewpoint is suggested, the total cost includes the cost of two viewpoints and is based on the joint number of detected peppers N_{JRP} and is calculated according to Equation 5.11.

$$TC = 2C_S + (N_P - N_{JRP})C_{UT} \tag{5.11}$$

5.3.4 Performance measures

The performance measures include:

Descriptive dataset measures

The measures include the average number of peppers in a scene $\hat{\lambda}$ as collected from the ground truth information, average occlusion level of a viewpoint $\overline{OL}$, average number of detected peppers $\overline{N_{RP}}$ (as detected by the manual annotators and the image analysis), and the accuracy measures of the target detection algorithm (precision and recall).

Suggested viewpoint evaluation measures

To evaluate the viewpoints suggested by the algorithm a number of measures are defined:

- **The number of targets prediction accuracy** measures the difference between the expected number of targets in the scene as predicted by Equation 5.5 and the actual number of targets in the scene N_P. The related measures include the average difference, root mean square error (RMSE), and normalized root mean square error (NRMSE). Similar measures are calculated for the difference between the expected number of additional targets to be sensed from an additional viewpoint as predicted by Equation 5.6 to the actual number of additional targets as sensed from the next viewpoint.
- **The detection increase (DI)** is defined as $N_{JRP} - N_{RP}$, and measures the number of additional targets detected when an additional viewpoint is added. The relative detection increase defined in Equation 5.12 measures the ratio between the detection increase and the ground truth information collected about the scene.

$$RDI = \frac{N_{JRP} - N_{RP}}{N_P} \tag{5.12}$$

- **The occlusion level reduction measure (OL_D)** is defined as the ratio of the occlusion level as calculated in Equation 5.2 between the original viewpoint image and the second

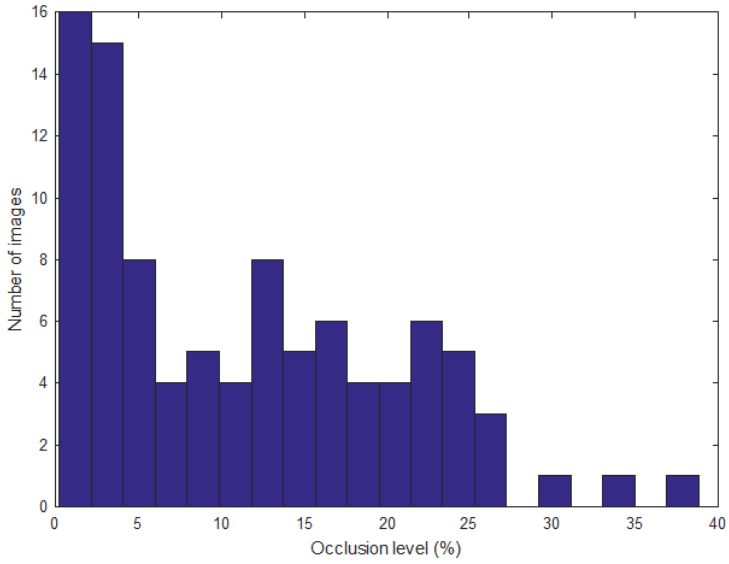


Figure 5.12: Occlusion level as detected in all images of the database.

viewpoint image. OL_D is measured only where an additional viewpoint is possible according to the decision process described in Section 5.2.4.

Sensing strategy evaluation

The relative total cost decrease (RTCD), calculated according to Equation 5.13 for each image taken, aims to find the relative decrease in the total cost as a result of the use of the dynamic sensing strategy proposed.

$$RTCD = 1 - \frac{TC_D}{TC_S} \tag{5.13}$$

where TC_S is the total cost as a result of the single viewpoint strategy and the TC_D is the total cost as a result of the dynamic sensing strategy. Additionally, the ratio of the number of cases in which an additional viewpoint is suggested is also presented.

5.3.5 Sensitivity analysis

Sensitivity analysis is performed for all defined measures (Section 5.3.4) as a function of changes in cost ratios and the number of detected peppers.

Cost ratios

Since the costs of sensing and undetected targets depend on a multitude of parameters (Section 5.2.5) [Elkoby, 2016; Elkoby et al., 2014], the analysis was conducted for the full range of ratios of C_{UT}/C_S to vary from 1 to infinity.

Number of detected peppers

To allow insights into the algorithm performance for both optimal and suboptimal detection algorithms, all measures are derived for both manual (human annotator) and automatically obtained (via image analysis) number of peppers.

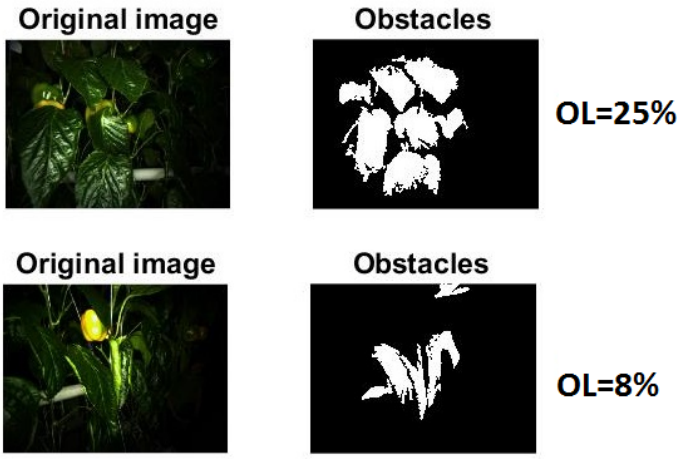


Figure 5.13: Images and detected obstacles. Top - high occlusion level. Bottom - low occlusion level.

5.4 Results

5.4.1 Descriptive dataset measures

The collected ground truth of the scenes in the greenhouse revealed 30 ripe peppers with an average of 3.75 and standard deviation(std.) of 1.4 peppers in a scene ($\hat{\lambda} = 3.75$). The average number of peppers in an image was 2.3 (std.=0.95) and 1.4 (std. 0.9) for the manual annotation ($\overline{N_{RP}^{M}}$) and automatic image analyses ($\overline{N_{RP}^{A}}$), respectively.
The median value of the detected occlusion level (OL) (Figure 5.12) of 10% was taken to separate into low and high occlusion levels, with examples presented in Figure 5.13.
The target detection algorithm had precision and recall rates of 0.95 and 0.63, respectively.

5.4.2 Parameters estimation

According to Equation 5.9, given the found $\overline{N_{RP}^{A}}$ and $\overline{N_{RP}^{M}}$, the image analysis detection probability and the manual evaluation detection probability are estimated as $\widehat{P_{D}^{A}} = 0.37$ and $\widehat{P_{D}^{M}} = 0.61$, respectively.

5.4.3 Suggested viewpoint evaluation measures

With a second viewpoint, detection rates increase by 19% from an average of 46% (95% confidence interval:39.0%-53.0%) detection to an average of 65% (95% confidence interval:57.6%-72.4%) detection. In 83.9% of the cases the next viewpoint resulted in lower occlusion levels. Separating the viewpoints into highly occluded ($OL > 10\%$) and low occluded viewpoints showed 84.3% reduction in occlusion level of the second viewpoint for viewpoints with high occlusion level, and an average 55% increase in occlusion level for low occluded viewpoints. Even though viewpoints with low occlusion level encountered a higher occlusion level in the second viewpoint, the additional viewpoint revealed additional targets at similar rates to the initial viewpoints with high occlusion level.

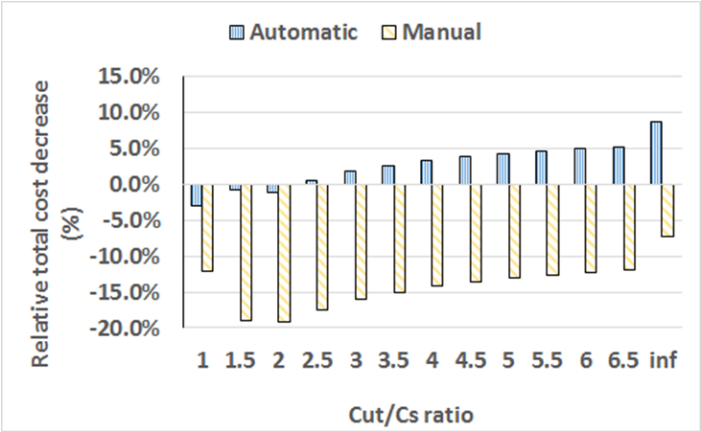


Figure 5.14: Relative total cost decrease.

The average difference between the expected number of targets as predicted in the scene according to Equation 5.5 to the actual number of peppers in the scene was 0.57 (RMSE=1.58, NRMSE=0.42). The average difference between the expected number of peppers to be detected from a second viewpoint according to Equation 5.6 and to the actual number of peppers detected was 0.89 (RMSE=1.03).

5.4.4 Sensing strategy evaluation

Results (Figure 5.14) reveal that if image analysis is non-optimal (i.e., when not all peppers in an image are revealed), a cost ratio higher than 3 yields a 5-10% cost decrease. The algorithm is not profitable when 100% detection is achieved in the first viewpoint; in these cases a single viewpoint should be employed since a second viewpoint is more costly than beneficial. These results reveal that the selected heuristic improves detection by combining viewpoints according to the recommended second viewpoint. For ideal manual detection the heuristic has failed to show any conditions in which additional detection overcomes the cost of an additional viewpoint.

The ratio of cases in which a second viewpoint is employed (Figure 5.15) reveal that for a ratio above 2 between costs all tested images in the database satisfied the condition defined in Equation 5.7. In this case a second viewpoint is suggested according to Section 5.2.2 in 58.3% of the cases. These results show that for a ratio above 2, if there is a viewpoint that is recommended by the heuristic it is always employed and therfore the detection remains constant.

5.5 Conclusions

The use of a dynamic sensing strategy, which chooses the next viewpoint if the cost/benefit associated with an additional viewpoint is beneficial, resulted in 19% increased detection rates with 5-10% decreased costs. These results were achieved for a small number of predefined specific viewpoints and occlusion levels. A larger set of viewpoints with higher resolutions in yaw, distance, and occlusion level might yield better results. Improved image analysis algorithms (e.g., [Zemmour et al., 2017; Vitzrabin & Edan, 2016a]) can further improve detection, however regardless, due to the inherent visibility limitation as noted in

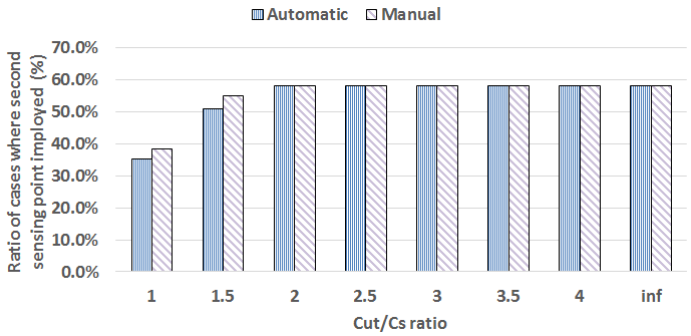


Figure 5.15: Ratio of cases in which a second viewpoint was employed.

the literature, adding a viewpoint is critical. For crops with inherently high visibility rate due to low occlusion levels (e.g., cherry tomatoes and grapes [Berenstein et al., 2010] - Figure 5.1), probably a single viewpoint strategy is sufficient to overcome the occlusion problem. This new concept of improving detection rates by dynamically adding a viewpoint is an important step towards advancing applicability of harvesting robots in high density crops. The proposed strategy was developed heuristically based on guidelines derived from viewpoints analyses for the specific crop.

To implement dynamic sensing for other crops the following main steps must be tailored to the specific application: 1) developing image processing algorithms for object and occlusion detection, 2) modeling the sensing costs assumptions, and 3) statistical analyses of different viewpoints to provide guidelines for developing heuristics related to the number of viewpoints and their combinations. For real-time implementation it is important to keep a simple decision process. However, future work should evaluate the benefit of finding the optimal location of an additional viewpoint versus a best-fit viewpoint selected using heuristics as in this work.

5.6 Research questions answered

The following research questions defined in Section 1.2 have been met:

RQ2.1: How can information content of a viewpoint should be measured? A measure of information content has been defined as the number of peppers detected compared to the number of expected peppers in the scene and the occlusion level.

RQ2.2: How to predict the overall number of peppers in the scene based on information gathered from one viewpoint? The number of peppers was modeled as a random Poisson variable. Therefore, the probability of detecting additional peppers was modeled as a conditional probability given the number of detected peppers in the first viewpoint.

RQ2.3: How to make a profitable decision on another viewpoint? Two costs were defined: the cost of loosing the undetected peppers and the cost of an additional viewpoint. The cost of loosing the undetected peppers depends on the number of peppers that were not detected. A comparison between the two costs leads to the decision if to add a viewpoint.

RQ2.4: Where should an additional viewpoint be placed? How is the location dependent on the information extracted from the first viewpoint? The features of the additional viewpoint are defined based on the occlusion level and the distance to targets extracted from the first viewpoint. Based on that data and a heuristic developed from the results of Chapter 4 the new viewpoint location is chosen from a list of predefined possible viewpoints. Future work should include an accurate calculation of the viewpoint.

Planning the sequence of tasks

6

> *The obvious rule of efficiency is you don't want to spend more time organizing than it's worth.*
>
> **— Daniel Levitin**
> (Cognitive psychologist, neuroscientist, writer.)

- Polina Kurtser, Yael Edan. "Planning the sequence of tasks for harvesting robots". Submitted to a robotics journal 2018.
- **Research objective RO3:** Compare different strategies of harvesting and sensing sequencing.

This chapter focuses on an algorithm for planning the sequence of tasks for a harvesting robot. The fruit targets are situated at unknown locations and must be detected by the robot through a sequence of sensing activities. Once the targets are detected, the robot must execute a harvest action at each target location. Planning the sequence of harvesting and sensing tasks is achieved using the traveling salesman paradigm by considering the costs of the sensing and harvesting actions along with the traveling times. The developed methodology is validated and evaluated in both laboratory and greenhouse conditions for a case study of a sweet pepper harvesting robot. The results indicate that planning the sequence of tasks for a sweet pepper harvesting robot can reduce travel cost on average by 12%. Incorporating the sensing operation in the planning sequence for fruit harvesting is a new approach in fruit harvesting robots and is important for cycle time reduction.

6.1 Introduction

Ongoing research is performed on harvesting robots, but still with no commercial success [Bac et al., 2014b; Hemming et al., 2014a]. The robotic harvesting cycle includes detecting the fruit, reaching the fruit, deciding whether it is ripe, and harvesting the fruit [Edan et al., 1991]. Major limitations are low detection rates—87% [Bac et al., 2014b] and low harvesting cycle times [Hemming et al., 2014a]. Most research to date focuses on improving detection algorithms [McCool et al., 2016; Sa et al., 2016; Bac et al., 2014b], and very little research is devoted to cycle time optimization [Mann et al., 2016; Zion et al., 2014; Edan et al., 1991] which is the main thrust of this chapter.

The complexity of the fruit detection task is due to the unstructured and dynamic nature of both the objects and the environment [Kapach et al., 2012]: fruits have a high inherent variability in size, shape, texture, and location; in addition, occlusion and variable illumination conditions significantly influence the detection performance. Previous research indicated that from a single viewpoint only 60% of the fruit are visible [Hemming et al., 2014b] and

that detection can be improved to 85-90% by combining multiple viewpoints. However, to provide better detectability it is important to select best-fit sensing points. Furthermore, since each sensing operation leads to additional travel and time costs it is important to plan these sensing tasks so as to reduce the overall cycle times.

Cycle times can be improved by optimizing the harvesting sequence [Edan et al., 1991]. The near-optimal harvesting sequence of fruits depends on the robot kinematics and the plant structure [Edan et al., 1991], and can be improved by applying solutions of the traveling salesman problem (TSP) to the fruit coordinates. Previous research of planning in harvesting robots assumed all fruit coordinates are known in advance (through previous detection). Since the TSP is NP hard [Lenstra & Kan, 1981], in most cases heuristics are used for large scale problems [Lawler, 1985]. Optimal solutions can be derived only for problems with few targets using exhaustive search [Laporte, 1992]. Algorithms to define the sequence of citrus harvesting were based on the TSP using the geodesic as the minimum path [Edan et al., 1991]; the sequence of melon picking for a multi-arm harvester was modeled as a task of coloring an interval graph based on GPS tagging of the fruit locations [Mann et al., 2016; Mann et al., 2014; Zion et al., 2014] before harvesting begins. In both works only the harvesting tasks were considered in the planning.

The traveling salesman problem, and the vehicle routing problem, in the context of agricultural applications has also been addressed for agricultural field logistics problems [Orfanou et al., 2013; Bochtis et al., 2010; Bochtis & Sørensen, 2010; Bochtis & Sørensen, 2009; Ali et al., 2009], where an action, such as watering or seeding, is required to be accomplished by multiple vehicles or humans or cooperation between the two. The required action is repetitive and static. Hence, the optimal plan was developed prior to actual traveling, relying on ideal knowledge of the targets and the cost of transportation between them.

In robotic harvesters to date, planning of the sensing operations has not been noted [McCool et al., 2016; Bontsema et al., 2015; Bac et al., 2014b]. Detection is conducted either from a fixed position or continuously using an "eye-in-hand" configuration [Barth et al., 2016; Bac et al., 2014b]. For harvesters that detected fruit a priori (e.g., [Bontsema et al., 2015; Bac et al., 2014b]), the harvesting sequence is usually defined using a heuristic sequence with some preset order such as distance to the sensing rig or along the lane [Bac et al., 2014b].

However, when detection is not performed a priori, continuous sensing is needed to reveal the fruit. This requires tying the harvesting sequence problem with planning the sensing actions to retrieve the target locations. Increasing the number of viewpoints increases cycle times and hence must be minimized.

Attempts to approach the question of planning of sensing actions, has been done before in other applications. This includes planning of sensor operations based on information content [Olawsky et al., 1993], and a stochastic planning approach, which models the task planning as a Markov Decision Process [Hansen, 1994].

The aim of this chapter is to develop a methodology for planning the sequence of tasks for a robotic harvester i.e., planning both the sensing and harvesting actions and motions to optimize the average harvesting time. To accommodate the possibility of other sensors that are not RGB (e.g. hyper spectral, thermal cameras, etc) in this chapter the term sensing point

(the physical location where sensing is performed) is substituting the viewpoint term used in previous chapters. The meaning remains the same - the pose (position & orientation) of the sensor, that needs to be achieved by the robotic manipulator. The value of this planning is evaluated by comparing the results to optimal harvesting, assuming all fruit locations are known a priori; and to common harvesting strategies to-date, which are mostly heuristic and do not employ any optimization method. The developed methodology is validated and evaluated in both laboratory and greenhouse conditions for a case study of a sweet pepper harvesting robot.

6.2 Formalization and algorithm

The methodology used to calculate the costs associated with traveling between the sensing points and targets to be harvested along with the sensing and harvesting operations is described. Then, the methodology of optimal sequence calculations is detailed given the locations of sensing points and sensed targets.

6.2.1 Formalization and notations

Given a 3D operational area A of size $L_R * L_C * L_D$, the following locations are defined:

- **Harvesting targets** $x_F = \{x_1 \dots x_{N_{FT}}\}$ are the N_{FT} fruit locations to be visited and serviced by performing a harvest action (e.g., grasping the fruit and cutting it).
- **Sensing points** $y = \{y_1 \dots y_{N_{SP}}\}$. The N_{SP} locations where sensing actions are performed for revealing unknown targets. Each sensing operation performed at point $y_i \in y$ reveals a subset of harvesting targets $x_{RF}(y_i) \subseteq x_F$.
- **Revealed harvesting targets** is the consolidation of the subsets of targets $x_{RF}(y_1) \dots x_{RF}(y_k)$ revealed in previous k sensing actions. If the robot have not performed any sensing operations yet then the revealed harvesting targets set is $\varnothing$.

The following costs are defined based on the robotic harvesting cycle:

- **Sensing cost** S_c. The cost of performing a sensing action, at sensing point y_i, for revealing undetected targets, i.e, the image processing time to detect the targets.
- **Traveling costs** $W_c(c_i, c_j)$. The cost of traveling between two points, a function of spatial coordinates (c_i can be a harvesting targets x_i or sensing points y_i) defining the cost of travel between them.
- **Harvesting cost** H_C. The cost of performing a harvesting action, at a target x_i, i.e., the time of the harvesting operation (e.g., grasp and cut, grasp and twist).

Both sensing cost S_C and harvesting cost H_C are considered constant since sensing and harvesting are repetitive tasks which are not dependent on the spatial locations of the sensing's or target's locations.

The travel sequence Seq_T is a list of ordered target and sensing coordinates that the robot travels through. The total traveling cost is defined as the path length in Equation 6.1.

$$T_{PL} = \sum_{i=1}^{N_{FT}+N_{SP}-1} W_c(Seq_T(i), Seq_T(i+1)) \tag{6.1}$$

The overall ***total travel cost*** T_c of a travel sequence Seq_T is defined by Equation 6.2, by adding the cost of travel to the cost of sensing and harvesting.

$$T_c = T_{PL} + H_C N_{TF} + S_C N_{SP} \tag{6.2}$$

For a constant number of targets N_{TF} and sensing points N_{SP} the total cost can be considered as the total travel cost T_C with an addition of a constant. The optimal travel sequence Seq_T^* is the sequence minimizing the total travel cost.

6.2.2 Target-sorting methods

Given $x_{RF}(y_i)$ revealed targets from sensing action y_i and the travel costs between them the following target-sorting methodologies were considered:

- **Optimal target-sorting method** sorts the targets according to the minimal distance between their spatial locations. This method solves the traveling salesman problem (TSP) for the given targets locations
- **Heuristic target-sorting method**. Sorts the coordinates of detected fruit in one dimension (e.g., near-to-far, right-to-left). This sequencing methodology does not aim to optimize the traveling cost, and is common in practice in many robotic harvesters.

An illustration of the different target-sorting methodologies is presented in Figure 6.1
Important to note that for the described application of a robotic harvester, the number of targets expected to be harvested from a single robotic platform location is $N_{FT} < 15$. This allows the optimal target-sorting method to solve the TSP problem using exhaustive search. Sweet peppers usually have an average of 3 peppers on a stem [Kurtser & Edan, 2018a]; this corresponds also to results by Sa et al.,[Sa et al., 2016] which revealed an average of 2-3 stems in an image. The database gathered in this chapter reports similar numbers. Other plants might have a larger amount of targets, and might require a sub-optimal TSP solution (Laporte, 1992; Lawler, 1985) of the traveling salesman problem. To avoid distraction of the main focus of the chapter, these cases are not covered and should be reviewed in the future.

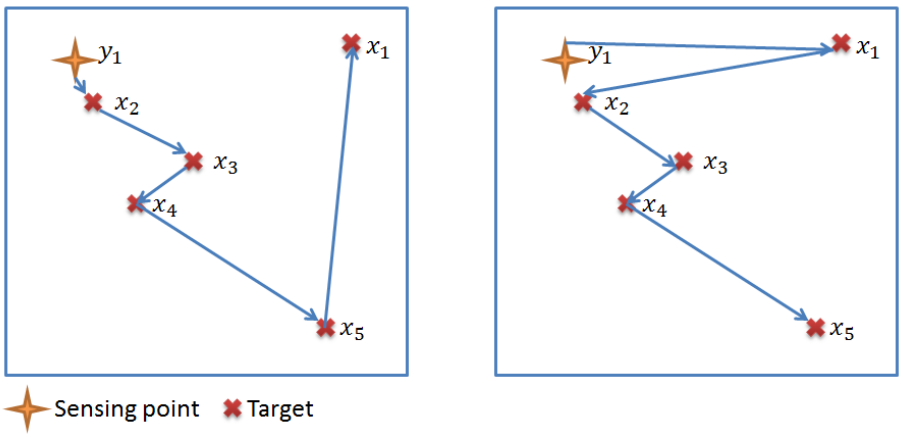


Figure 6.1: Illustration of target-sorting methods for sensing point y_1 and the revealed targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$. Left: optimal target sorting method. Right: Heuristic target sorting method (top-down).

6.2.3 Sensing methods

To obtain global full information about the targets locations several sensing operations are required. It is assumed that all sensing points' locations are set a priori based on optimal viewpoint analyses. The sensing method defines when the sensing operation will be performed within the travel sequence. The following sensing methods were compared (Figure 6.2):

- **"A priori-sensing"** – "Sense-all-first followed by harvest-all". In this case all sensing points are performed a priori before the harvesting sequence is derived, to obtain locations of all targets and perform a global optimum calculation of the targets harvesting sequence.
- **"Batch-sensing"** – "look and move". Each sensing point is followed by harvest of all targets revealed as a result of the previous sensing action. The next sensing point is selected only after the harvesting of all targets revealed is completed.
- **"Sensing-in-harvest"** – "sensing included in harvesting sequence". The sensing points are included in the list of locations to be visited. When the robot travels between target locations, and reaches a sensing location, it performs a sensing action. The new revealed targets are now integrated into the whole list of targets and a new optimized sequence is calculated.

In the example presented in Figure 6.2, the "*a-priori sensing*" is where y_1 and y_2 are traveled to first and then all the revealed targets $x_I = \{x_1 \dots x_8\}$ are sorted top-down. The batch sensing includes traveling to y_1 where targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$ are revealed, and then sorting $x_{RF}(y_1)$ in a top-down order, then y_2 is visited where $x_{RF}(y_2) = \{x_6 \dots x_8\}$ are revealed and sorted in a top-down order. Finally, the sensing-in-harvest includes visit to y_1 where targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$ are revealed, followed by sorting $x_{RF}(y_1)$ as well as y_2 is a top-down order. In this case y_2 is visited after visiting targets x_1 and x_2 where $x_{RF}(y_2) = \{x_6 \dots x_8\}$ are revealed as a result targets $\{x_3 \dots x_8\}$ are re-ordered in a top-down order.
It is important to note that the "*a priori-sensing*" method cannot be used as the optimal

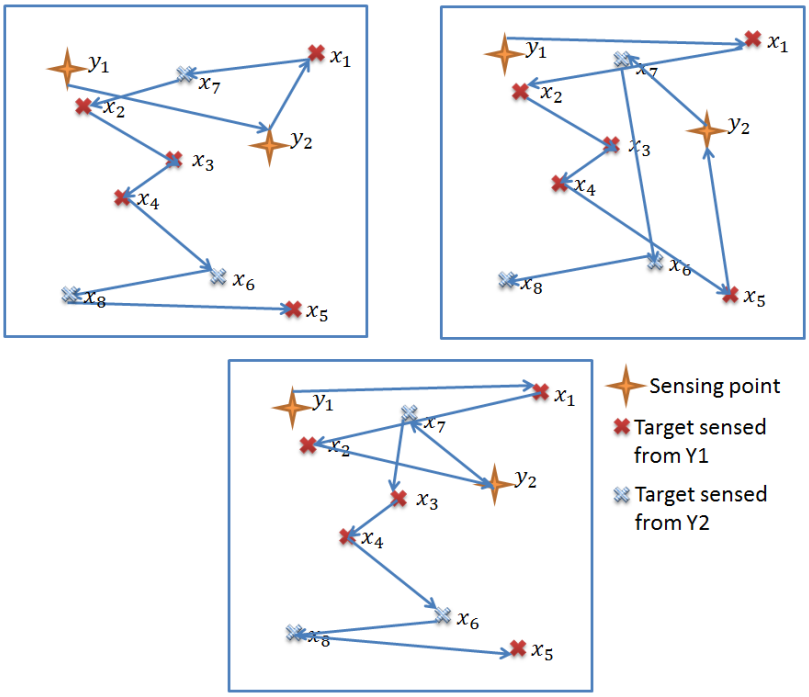


Figure 6.2: Illustration of sensing methods for two sensing points y_1 and y_2 and the revealed targets $x_{RF}(y_1) = \{x_1 \dots x_5\}$, $x_{RF}(y_2) = \{x_6, x_7, x_8\}$. A top-down target-sorting is used. Top-left: a-priori sensing. Top-right: batch sensing, Bottom: Sensing in harvest.

benchmark since it is not necessarily the optimal solution. The optimal sequence is highly dependent on the unknown fruit distribution and on the sensing points' locations. It is also important to note that "a priori-sensing" can be performed by a different system/robot. In this case the costs of travel to the sensing points might be minimized by having the system designed for overview points locations. However the cost of an extra system in place and the registration errors derived should be considered.

6.2.4 Harvesting targets location and sensing data

The expected locations of the targets in space are highly dependent on the application. While in some applications, the locations of the targets can be assumed to be uniformly distributed, in agricultural applications some areas have more probability for target occurrence than others due to the biological nature of the problem (i.e., fruits will grow on a stem, in a certain height window, with a certain growth distribution). The target location is described as a random variable drawn from a 3D ***target locations probability function*** $P_T(c)$ defining the probability of the target to be present at the coordinate $c(x, y, z) \in A$. Given target x_i, a ***revelation probability*** P_{R_j} defines the probability of the target to be revealed at a sensing point y_j. This generates a target-sensing pairs where a sensing action at sensing point y_j reveals a subset of targets $x_{RF_j} \in x_F$.

6.2.5 Traveling costs definition

The traveling cost $W_c(c_i, c_j)$ between two spatial coordinates c_i, c_j (targets or sensing points) depends on the traveling agent degrees of freedom, on the path, and its motion profile. The

cost is defined as the time of travel between two coordinates by a robotic arm. The general notation presented in this research can be generalized to other traveling costs as well, such as energy consumption and traveled distance. The following methods for traveling costs estimation are used in this chapter:

- **Time of travel with a robotic arm performing point-to-point motion (time matrix)**. In this case the robotic arm performs a point-to-point motion between two coordinates according to the intrinsic velocity profile of the robotic arm. A point-to-point motion is a path minimizing the squared sum of angular distances of the joints required to travel between c_i and c_j. The actual times of travel are registered and are used as the traveling cost.
- **Weighted sum of angular distances of the RRT connect path solution**. In this case, the path between each two points is calculated using the RRT connect algorithm [Kuffner & LaValle, 2000] for path solution. The path length is then defined as the root of the sum of square angular distances of the joints required to travel between c_i and c_j weighted by joint angular speed ω.

The time travel matrix approach does not make a priori assumptions regarding the path planning algorithm or internal joint speeds of the robot. The weighted sum approach assumes the robot path planner is RRT connected and that the angular speed of the joints is known. The time travel matrix does not assume symmetry, and allows the cost of travel between c_i and c_j to be different from the cost of travel from c_j to c_i.

6.3 Methods

To validate and evaluate the proposed methodology in both laboratory and greenhouse conditions, the following experimental protocols were applied:

1. **Datasets generation and acquisition protocols** where target points and sensing points are registered in both a simulative environment and greenhouse environment.
2. **Target-Sensing dataset**. The datasets are processed into a list of sensing points and the resulting revealed targets in each scene as a result of the revelation probability or actual information gathered in the greenhouse.
3. **Acquisition of traveling cost between all points**. Acquisition of the traveling costs between each two points according to the defined measures of $W_c(c_i, c_j)$.
4. **Application of the sensing and sorting methodologies and performance measures calculation**. Given the targets and the travel costs the different sensing and sequencing methodologies are applied. For each of the proposed sensing and sequencing methodologies a set of performance measures is calculated to evaluate the best sensing and sequencing methods given the revelation probability, traveling cost, and experimental hardware setup.

6.3.1 Datasets generation and acquisition protocols

Two datasets were used – an artificial database (DB#1) and a greenhouse database (DB#2). The greenhouse database (DB#2) consists of known locations extracted from greenhouse

Figure 6.3: Robot acquisition of target locations and sensing points in the greenhouse as well as axis origin and orientation in the robotic base.

data. 3D locations are based on data extracted from experiments performed in greenhouse conditions, generating a 3D database of actual locations, retaining a 10^{-2} decimal accuracy. The greenhouse dataset (DB#2) was acquired in April, 2017 in a research greenhouse at Sint-Katelijne-Waver, Belgium, using a 6 degree of freedom manipulator (Fanuc LR Mate 200iD). The robot was equipped with an iDS Ui-5250RE RGB camera (Figure 6.3) to automatically acquire images from 12 viewpoints described in Table 6.2 with artificial illumination conditions. It includes 3D locations of 12 peppers in 2 scenes as registered early in the harvesting season according to the protocol described in [Barth et al., 2016].
To generate the artificial database (DB#1) a **theoretical distribution ($P_T(c)$)** was used. $P_T(c)$ is defined based on distributions assumed to be relevant to the application. In pepper harvesting, the targets are expected within a vicinity of the stem in a finite range in height. In commercial greenhouses the stems are placed at a constant distance from one another. Given $P_T(c)$ the DB#1 is generated by randomly drawing target coordinates according to this distribution, retaining a 10^{-2} decimal accuracy. The artificial dataset (DB#1) includes 12 artificially generated scenes. Each scene includes 7 targets and 2 sensing points with locations drawn from the location probability functions described in Table 6.1 according to the axis describe in Figure 6.3. All randomly selected locations used are rounded at $10^{-2}cm$ accuracy. The locations were separated into three groups:

- **Full 3D uniform distribution**. In this case, the full area reachable by the robot was used, forming a cube centered around the robot, eliminating the area too close to the robot and therefore unreachable. Both targets and sensing points were chosen from that area.
- **One side 3D uniform distribution**. In this case, limiting the X-axis to only positive values allows the evaluation of the case commonly applied in the greenhouse where only one side of the lane is harvested at a time. Therefore, only one side of the robot is sensed and harvested. Both targets and sensing points are chosen from the same area.
- **Greenhouse distribution**. In this case one-sided X-axis limited area is split into two narrow zones in the Y-axis, each 5 cm wide. Sensing points are taken closer to the robot than the targets and therefore the sensing points were limited by X.

Table 6.1: DB#1 – Generated scenes according to the axis in Figure 6.3.

Scene	Location type	Coordinate	Distribution $P_T(c)$
1-4 Full 3D	Target	X	$X \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Y	$Y \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Z	$Z \sim U(0.3,0.8)$
	Sensing	X	$X \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Y	$Y \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Z	$Z \sim U(0.3,0.8)$
5-8 One side	Target	X	$X \sim U(0.3,0.6)$
		Y	$Y \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Z	$Z \sim U(0.3,0.8)$
	Sensing	X	$X \sim U(0.3,0.6)$
		Y	$Y \sim U(-0.3,-0.6)\&U(0.3,0.6)$
		Z	$Z \sim U(0.3,0.8)$
9-12 Greenhouse	Target	X	$X \sim U(0.4,0.6)$
		Y	$Y \sim U(0.33,0.38)\&U(0.53,0.58)$
		Z	$Z \sim U(0.3,0.8)$
	Sensing	X	$X \sim U(0.3,0.4)$
		Y	$Y \sim U(0.3,0.6)$
		Z	$Z \sim U(0.3,0.8)$

Table 6.2: DB#2 - 12 sensing points' description of locations and orientation.

Sensing point	Distance (mm)	Tilt (degrees)	Azimuth (degrees)
1	19	20	0
2	19	-20	0
3	8	-20	0
4	19	10	-50
5	10	10	-50
6	10	0	-50
7	10	20	0
8	9	0	0
9	7	20	20
10	7	0	20
11	19	20	20
12	17	0	0

6.3.2 Target-Sensing dataset

This process generates a list of target-sensing sets where each sensing point has a subset of targets that are revealed as a result of performing a sensing action at that spatial point. As mention in Section 6.2.4, the ***revelation probability*** P_{R_j} defines the probability of the target to be revealed at a sensing point y_j. In this protocol this probability is defined in two ways

- **Constant probability**. Previous research [Hemming et al., 2014b] has shown that 40-60% of peppers are revealed from a single sensing point. Therefore, as part of the simulation protocol a constant revelation probability $P_{R_j} = [0.4, 0.5, 0.6]$ is defined at each sensing point, in a way that all simulated targets have P_{R_J} probability to be revealed from any sensing point. Meaning, for each sensing point y_j, n random variables $I_{i,j} \sim Bernouly(P_{R_J}) \quad \forall \quad i = 1 \dots n$ are drawn. If $I_{ij} = 1$ then target x_i is revealed from sensing point y_j else it is not. By doing so, independence between the detectability of the targets on other targets is assumed to maintain a constant

Figure 6.4: Detectability rate example marking Left: 50%; Right: 25%.

detectability rate. As a result DB#1 is processed into a list of simulated sensing points and target locations revealed from the sensing points.

- **Known sensing point information from greenhouse data**. The greenhouse dataset DB#2 also includes sensing points and the detectability of the targets from each of them. This data is derived from the manual data tagging precedure (Chapter 5). The detectability of a target is marked on a scale of 0% to 100% with 25% increments, representing the revealed part of the pepper (Figure 6.4). A combination of two sensing points is chosen according to the following criteria: (i) the combination of sensing points must provide full detectability of all peppers available in the scene; and (ii) the combination must provide the maximum sum of joint detectability of all peppers among the combinations complying with the first criteria. This information is processed to derive the two best sensing points and the list of targets revealed from each of them.

6.3.3 Acquisition of traveling cost between all points

To derive the traveling costs described in Section 6.2.5 the following procedures were implemented.

A 7-DOF Kuka LWR4 was placed in a simulated environment using the V-Rep version 3.4 simulation software. The simulated targets and sensing points of DB#1 were placed in the scene as shown in Figure 6.5. In the simulated environment the robotic manipulator moved from one point to another using the solution of the RRT connect algorithm provided by the OMPL software package. At each point the cost of travel to all unvisited points was calculated and registered according to the weighted sum of angular distances of the path with joint angular velocities of $\{1,1,1,1,1,1,1\}$.

The extracted 3D locations of targets and sensing points from greenhouse DB#2 were implemented in a harvesting sequence of a 7-DOF collaborative robotic arm Kuka LBR iiwa in laboratorial conditions (Figure 6.6). The robotic arm moved from each registered point to another and the time of travel was logged. The time assumed needed for harvesting a pepper was assigned to be 3 sec and the time of sensing processing at each viewpoint was assigned to be 1 sec corresponding to required time operations for successful robotic harvesting [Elkoby et al., 2014].

These procedures resulted in a travel cost matrix with distances of travel between each two points in the scene for both DB#1 and DB#2.

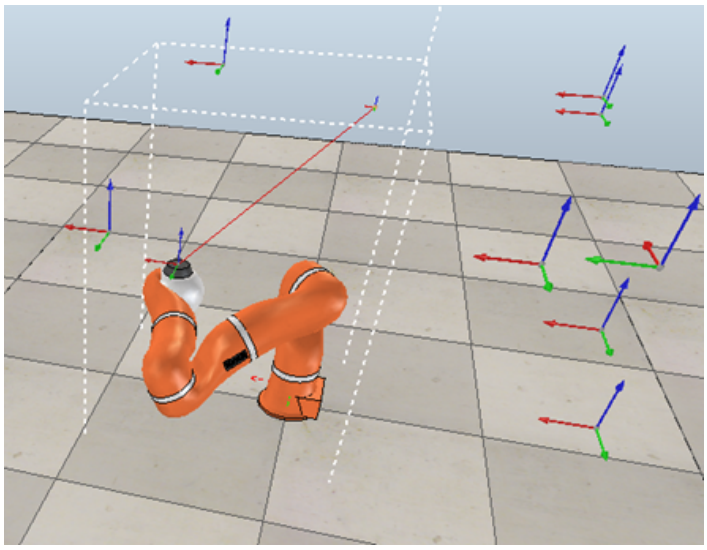


Figure 6.5: Simulated environment of the Kuka LWR4 robot and the targets.

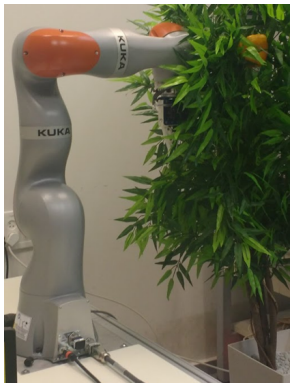


Figure 6.6: Laboratory reconstruction of the greenhouse conditions on a KUKA LBR iiwa.

6.3.4 Application of the sensing and sequencing methodologies and performance measures calculation

Given the traveling cost matrix of a scene all combinations of sequencing and sensing methodologies were applied both in simulated and laboratory conditions. The heuristic sequencing method included sorting the location in different common strategies: (i) right-to-left; (ii) up-to-down; (iii) near-to-far. For each generated sequence the following measures were calculated:

- **Average travel cost for fruit ($\bar{T_c}$).** Given sequence travel cost T_C as described in Equation 6.1 and N_{RT} number of jointly revealed targets as a result of all sensing operations, the average travel cost for fruit is defined as ($\bar{T_c} = T_C / N_{RT}$).
- **Joint detectability**. Defined as the ratio of revealed targets as a result of all sensing operations, and the overall number of targets in the scene. For example, let us assume a scene where 5 peppers ($x_1 ... x_5$) are to be harvested. Let's assume from the first viewpoints (y_1) only x_1, x_3 and x_4 has been reveled. From the second viewpoint (y_2)

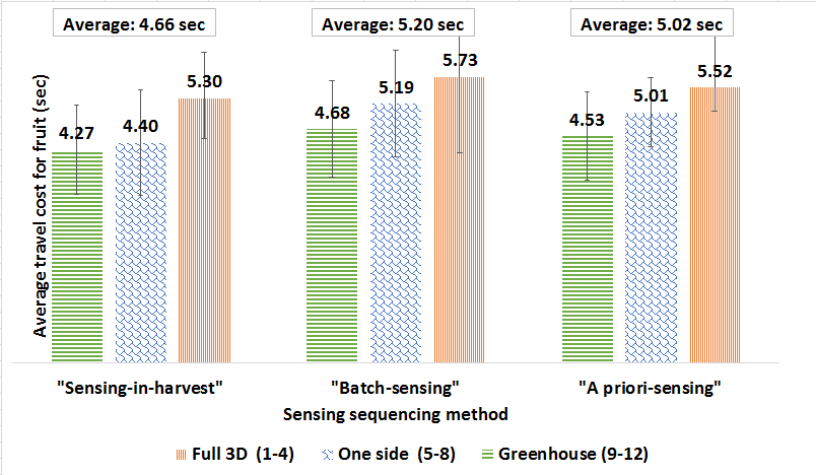


Figure 6.7: Average travel cost per fruit as a function of sensing sequencing methodology and location probability function.

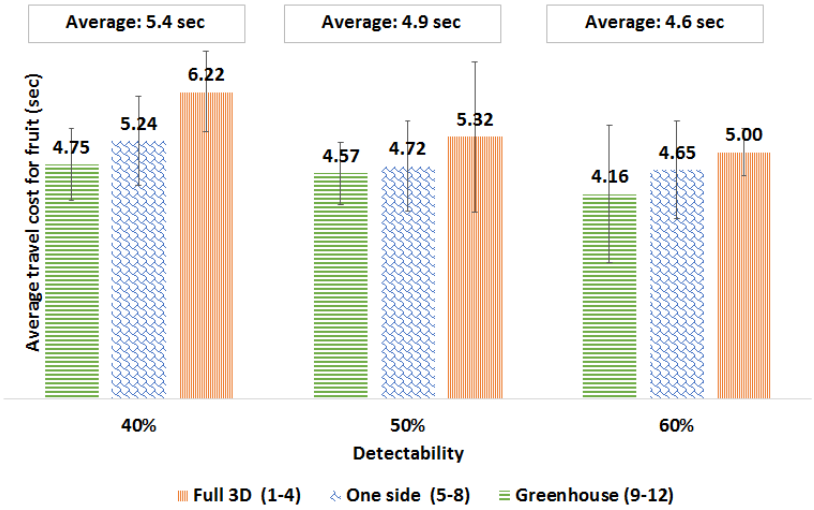


Figure 6.8: Average travel cost per fruit as a function of revelation probability and location probability function.

x_1 and x_2 are revealed. The detectability from the first viewpoint is $3/5$ and from the second viewpoint is $2/5$ while the joint detectability from the two viewpoints is $4/5$.

The simulated procedure was performed once for each combination of sensing sequencing methods, while in laboratory conditions each combination was performed three times for each sensing sequence and sequencing method.

6.4 Results

The results of the average travel cost for fruit calculated as a weighted sum of angular distances of artificial DB#1 yield an average 8% and 12% decrease in travel cost for "*sensing in harvest*" sequence compared to "*a priori-sensing*" and "*batch-sensing*" respectively (Figure 6.7). The analysis of the weighted sum of angular distances' travel cost as a function of sensing sequencing methodologies under different target locations probability functions is presented in Figure 6.7. Results indicate that the "*sensing-in-harvest*" sequence is the

Table 6.3: Average decrease in travel cost for "sensing-in-harvest" sensing methodology compared to other sensing methods as function of target probability functions.

Target locations probability functions	Average decrease in travel cost	
	"a-priori-sensing"	"batch-sensing"
Full 3D	4%	8%
One side	13%	17%
Greenhouse	6%	10%
Combined	8%	12%

sequencing methodology that yields best results (minimum travel cost) for all tested location probability functions, with a highest decrease of 17% for a one sided location probability function as compared to "*batch-sensing*" (Table 6.3).

The results of the same average travel cost as a function of revelation probability P_{R_j} resulted in an average 6% cost increase for 50% revelation probability compared to 60% revelation probability, and 17% cost increase for 40% revelation probability compared to 60% (Figure 6.8). The average travel cost as a function of revelation probability for different location probability functions is presented in Figure 6.4. Results indicated maximum 24% reduction in travel cost with higher revelation probability (Table 6.4).

The results of the same average travel cost as a function of revelation probability for different sensing sequencing methodologies are presented in Figure 6.9. Average decrease in travel cost is up to 18% (Table 6.5). A more in depth analysis of this relation is presented in Figure 6.10, indicating the relation between the same average traveling cost and the joint detectability as defined in Section 6.3.4. Results show a linear negative relation where the higher the joint detectability the lower the average cost of travel. The "*sensing-in-harvesting*" method consistently results in better performance at all detectability levels, followed by the "*a priori-sensing*" and the "*batch-sensing*".

The total travel costs T_c obtained using the greenhouse data-based experimental protocol as a function of the sorting and sensing methodologies are presented in Figure 6.11. "*batch-sensing*" resulted in worst performance in all sorting methodologies, with an increase of up to 9% in overall traveling costs compared to the other sensing methodologies. "*A priori-sensing*" resulted in best or equivalent results to "*sensing-in-harvest*". "*Sensing-in-harvest*" resulted in similar performance to "*a priori-sensing*" cases where the heuristic sorting methods yield optimal sequencing.

When the sorting was performed according to a heuristic (right-to-left or bottom-up), performing "*sensing-in-harvest*" caused an increase of up to 6% in travel time compared to the other methods. This implies that for a non-optimal sorting heuristic, "*a priori-sensing*" still yields the best results compared to sequencing the sensing by using a non-optimal sorting heuristic. However, when applying "*a priori-sensing*" in a right-to-left sequencing approach yields similar results (only 0.6% increase), as compared to the time matrix sequencing. This can be explained by the growth model in the greenhouse where peppers grow at similar heights, in clusters, and therefore harvesting in a sequential manner stem-by-stem often yields the same harvesting sequence as the optimal sequencing.

Table 6.4: Average decrease in travel cost 60% revelation probability compared to lower revelation probability as function of target probability functions.

Target locations probability functions	Average decrease in travel cost	
	$P_{R_J} = 50\%$	$P_{R_J} = 40\%$
Full 3D	6%	24%
One side	2%	12%
Greenhouse	10%	14%
Combined	6%	17%

Table 6.5: Average decrease in travel cost for "sensing-in-harvest" sensing methodology compared to other sensing methods as function of revelation probability.

Revelation probability P_{R_J}	Average decrease in travel cost	
	"a priori-sensing"	"batch-sensing"
60%	5%	8%
50%	5%	9%
40%	12%	18%

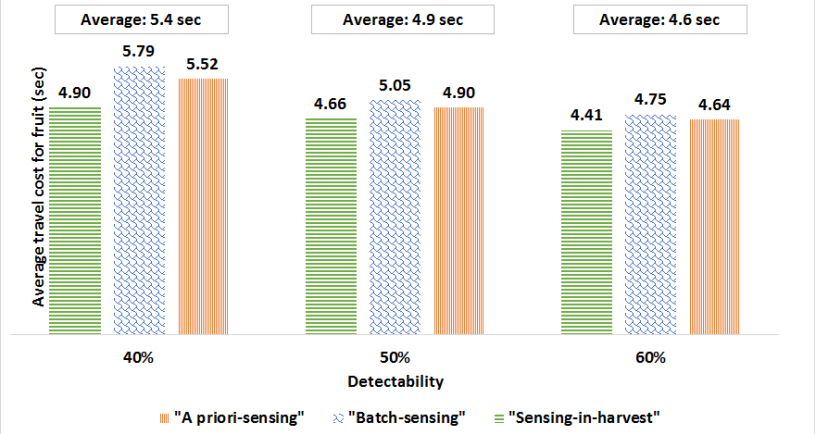


Figure 6.9: Average travel cost per fruit as a function of revelation probability and sensing methodology.

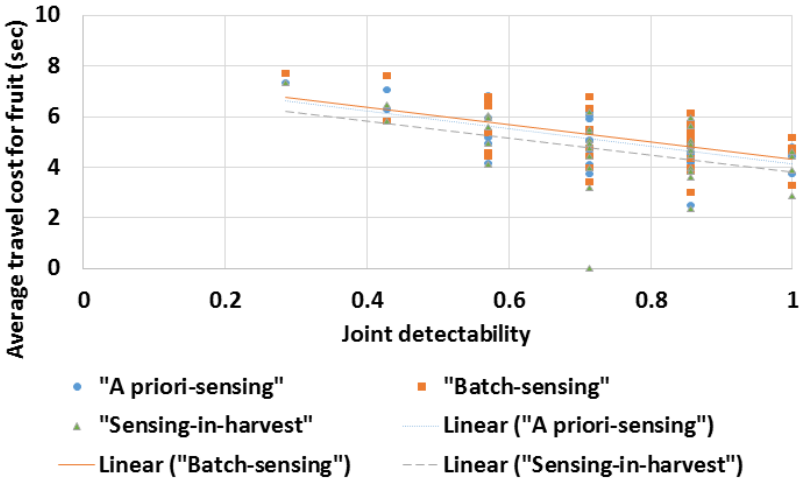


Figure 6.10: Average travel cost per fruit as s function of sensing methodology and joint detectability.

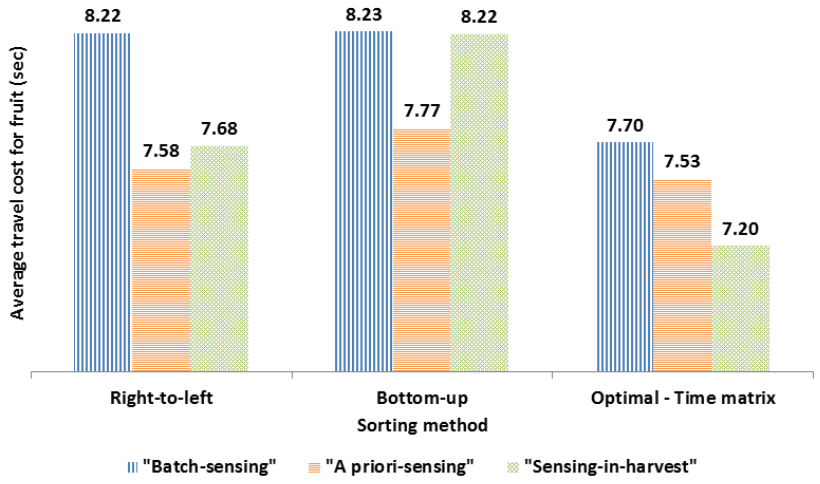


Figure 6.11: Total cost as a function of sequencing and sensing methodology.

6.5 Conclusions

A method to plan the sequence of tasks for a harvesting robot has been presented and defines the task sequence considering both harvesting and sensing tasks. This is the first time sensing operations have been considered in the planning of tasks for a harvesting robot.

For the case study of a sweet pepper harvesting robot, planning the sensing operations combined with planning of harvesting motions ("*sensing-in-harvest*") reduces overall travel costs by 12% on average, compared to the currently widely used approach of "*batch-sensing*" where harvesting is performed immediately after sensing operations.

The results of the simulation-based data show that "*sensing-in-harvest*" sensing sequence performs best to minimize average travel cost. This difference is particularly dominant as the variability of the fruits location probability function increases. When the sensing points were different in their location probability function, the difference between "*a priori-sensing*" and "*sensing-in-harvest*" was reduced even more due to the tendency of the "*sensing-in-harvest*" solution to group the sensing points together, similar to the "*a priori-sensing*" methodology. This becomes even more vivid in the result of the greenhouse data. The results indicate that "*a priori- sensing*" yields significantly better results than "*batch-sensing*", but close or equivalent results to the "*sensing-in-harvest*" sequence. However, this conclusion is limited by the assumption that once a target has been detected it will remain static in its location until harvested. Since the environment is highly flexible and dense the fruits can move from their initial location while the robot harvests; hence, the "*look and move*" approach is currently the dominant approach. Nevertheless, the estimated decrease of travel cost as a result of "*batch-sensing*" is high, and should be considered in further developments of harvesting robots to reduce cycle times.

6.6 Research questions answered

The following research questions defined in Section 1.2 have been met:

RQ3.1: How to calculate the traveling distance/cost function given a robot configuration? Different cost functions were evaluated including distance calculated using

a path planner and an empirically derived distance matrix, measured from actual robotic motions.

RQ3.2: What is the travel time decrease if optimization is introduced in the described conditions with full a-priori knowledge of target locations? As expected the travel time decreased with introduction of sequencing when a proper cost function was introduced for both simulative and laboratorial conditions.

RQ3.3: How unknown locations should be treated in planning of a harvesting sequence? Inclusion of the additional sensing points into the overall traveling sequence and readjusting the plan when new target has been detected ("*sensing-in-harvest*") has revealed reduction in travel time.

7 Evaluation of approach strategies for harvesting robots

> *I think that if your approach is one where you don't want to alienate anybody, you're going to have to soften the viewpoint or the information that you're offering to such an extent that it doesn't have the power to make any difference. You have to take that risk.*
>
> **— Eddie Vedder**
> (American musician, multi-instrumentalist and singer-songwriter)

- Published in: Ringdahl, Ola, Polina Kurtser, & Yael Edan (2018). "Evaluation of approach strategies for harvesting robots: Case study of sweet pepper harvesting". Journal of Intelligent and Robotic Systems,Springer,pp.1-11.
- Based on the publications: Ringdahl, Ola, Polina Kurtser, & Yael Edan (2017). "Strategies for selecting best approach direction for a sweet-pepper harvesting robot". Towards Autonomous Robotic Systems: 18th Annual Conference. Guildford, UK: Springer, pp. 516–525.
- **Research objective RO4:** Compare different approach strategies.

Robotic harvesters that use visual servoing must choose the best direction from which to approach the fruit to minimize occlusion and avoid obstacles that might interfere with the detection along the approach. This work proposes different approach strategies, compares them in terms of cycle times, and presents a failure analysis methodology of the different approach strategies. The different approach strategies are: in-field assessment by human observers, evaluation based on an overview image using advanced algorithms or remote human observers, or attempting multiple approach directions until the fruit is successfully reached. In the latter approach, each attempt costs time, which is a major bottleneck in bringing harvesting robots into the market. Alternatively, a single approach strategy that only attempts one direction can be applied if the best approach direction is known a-priori. The different approach strategies were evaluated for a case study of sweet pepper harvesting, in laboratorial and greenhouse conditions.
The first experiment, conducted in a commercial greenhouse, revealed that the fruit approach cycle time increased 8% and 116% for reachable and unreachable fruits respectively when the multiple approach strategy was applied, compared to the single

approach strategy.
The second experiment measured human observers' ability to provide insights to approach directions based on overview images taken in both greenhouse and laboratorial conditions. Results revealed that human observers are accurate in detecting unapproachable directions while they tend to miss approachable directions.
By detecting fruits that are unreachable (via automatic algorithms or human operators), harvesting cycle times can be significantly shortened leading to improved commercial feasibility of harvesting robots.

7.1 Introduction

Due to the lack of skilled workforce and increasing labour costs, advanced automation is required for greenhouse production systems [Comba et al., 2010]. The development of autonomous robots [Mann et al., 2016; Bontsema et al., 2015; Urrea & Muñoz, 2015; Edan et al., 1993] for agriculture aims to fulfill that requirement. Despite ongoing research on harvesting robots, harvesting robtos to date has yet penetrated the market and there are not in commercial use [Bac, 2015; Bac et al., 2014b]. Robotic harvesting includes several tasks: detecting the fruit, approaching it, deciding whether the fruit is ripe, and finally grasping the fruit and detaching it from the stem [Edan et al., 1991]. The steps are further described below.

Detection is considered to be one of the major limitations preventing commercialization of autonomous harvesting robots today with state of the art detection rate limited at 85% [Bac et al., 2014b]. A major problem is the unstructured and dynamic nature of the agricultural environments [Kapach et al., 2012]: fruits have a high inherent variability in size, shape, texture, and location; in addition, variable illumination conditions and occlusion significantly influence the detection performance. Significant research have been focused on developing detection algorithms [Gongal et al., 2015; Bac et al., 2014b; Hemming et al., 2014b]. Variable illumination conditions have been overcome using different techniques such as adaptive thresholding (e.g. [Vitzrabin & Edan, 2016a]), adding controlled illumination (e.g. [Font et al., 2014]), applying high dynamic range cameras [Suh et al., 2018], and color modification [Tang et al., 2016]. Attempts to deal with occlusion have previously been made using hyperspectral imaging [Lass & Prather, 2004]. However, due to the high cost as well as the weight of hyperspectral cameras, RGB cameras have become the most commonly used sensor since fruits, particularly ripe, tend to be different in color than the background [Bac et al., 2014b]. Additional solutions include mechanical removal of occlusion in a temporal manner using air blowing [Yoshida et al., 1985], or permanently by pruning leaves. Both methods, from internal conversations with growers and experts in the field, cause disagreement in the growers community due to the possible impact on yield and damage to the plant. Pruning of leaves also involves a lot of manual work which will impact the economical feasibility of a harvesting robot.

Detecting the fruit is typically done from an overview waypoint W_0 where several plants are visible [Bontsema et al., 2015] (Figures 7.1 & 7.5). This can be done either by using external static sensors or eye-in-hand cameras mounted on a robotic manipulator [Bontsema et al.,

2015] with different algorithms [Zemmour et al., 2017; McCool et al., 2016; Sa et al., 2016; Vitzrabin & Edan, 2016a; Bac et al., 2014b; Kapach et al., 2012]. Viewpoints analyses in harvesting robotics indicate that only 60% of the fruit can be detected from a single detection direction [Hemming et al., 2014b]. Improved performance can be achieved by including a human in the interpretation of the visual information received from the robot sensors [Berenstein & Edan, 2012; Bechar & Edan, 2003].

Approaching the fruit when eye-in-hand cameras are used is often divided into two steps [Barth et al., 2016; Han et al., 2012]. The first step involves moving from the overview waypoint W_0 to an approach waypoint W_1 where a single fruit is centered in the image. This waypoint is identified using the approach direction θ_1', defined as the angle the robot should use to approach the fruit from, and distance d to the fruit, as seen in Figure 7.1. The second step uses visual servoing [Barth et al., 2016; Chang, 2007] to move towards the fruit until it is reached. The first step is not always used [Kitamura & Oka, 2005], but can be introduced to shorten the time it takes to reach the fruit since the distance needed for visual servoing will be shorter, while the second step is used to continuously refine the fruit position detected from the overview image. The first step is important so as to reduce the long overall harvesting cycle time which is another major limitation preventing commercializing harvesting robots [Bac et al., 2014b]. Planning a path towards the fruit needs to take into account plant stems and other obstacles to prevent harming the vegetation [Bac et al., 2016; Bac, 2015; Bac et al., 2014a]. A methodology to derive reaching cones for agricultural environment characterization was developed by Bloch et al.,[Bloch, 2017]. The work is a preliminary method to describe obstacle-free areas for the robotś motion however is not applicable for real-time approaching of a fruit. When the fruit has been reached, or sometimes on the approach towards the fruit, the maturity of the fruit must be evaluated to determine if it should be harvested or not [Harel et al., 2016]. If it is determined to be **ripe** enough, the fruit must then be **grasped**. The accurate grasping of a fruit is a difficult problem due to the limitations of available robotic grippers and the inherent difficulties of grasp planning [Eizicovits & Berman, 2014; Rosenbaum et al., 2006]. Eizicovits and Berman [Eizicovits & Berman, 2014] developed geometry-based grasp quality measures based on 3D point cloud to determine the best grasping pose of different objects, including sweet peppers. This kind of solution depends on detailed 3D sensor information of the object [Eizicovits et al., 2016; Lehnert et al., 2016] which is very difficult to achieve in dense greenhouse environments. Obtaining this information in enough detail prolongs the harvesting cycle. Some gripper solutions that do not need an accurate grasping pose have been reported, but currently the harvesting success is limited [Bac et al., 2017]. Once the fruit has been grasped, it must be **detached** without damaging the plant or fruit. This operation is fruit dependent (e.g., for sweet peppers the detaching is performed by cutting the peduncle of the pepper; for apples a twist and snap operation is needed).

Given the dense environment and the continuous detection required when using visual servoing, there is a high risk of losing the fruit along the approach due to occlusion, regardless of the detection algorithm used. To reduce the risk of losing a fruit due to occlusion it is important to approach it from a waypoint from where the fruit is not occluded by leaves

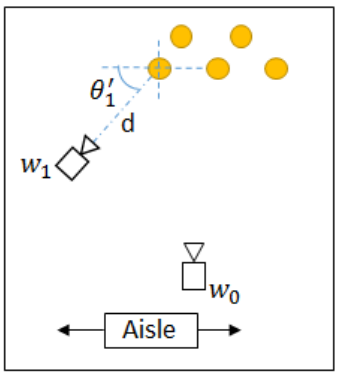


Figure 7.1: Overview waypoint W_0 (camera facing front direction) and approach waypoint W_1, identified using the distance to fruit d and approach direction θ_1'.

and other obstacles along the approach. An *approach strategy* is the method of finding an approach waypoint from which visual servoing will not lose the fruit due to occlusion. The aim of this chapter is to propose different approach strategies and compare them in terms of cycle times and success rates and identify failure causes. To focus on the approach task, standard color segmentation algorithms for detecting fruits and ripeness are used while limitations of grasping and detaching the fruits are not regarded. The work is demonstrated for a case study of sweet pepper (*Capsicum annuum*) harvesting in a research greenhouse as part of the Horizon 2020 EU SWEEPER project (G.A. 644313)[1] using a robotic manipulator equipped with an eye-in-hand camera. Previous limited work in lab conditions revealed that the choice of approach strategy influenced cycle time up to 40–45% but did not influence the success rate, which was 100% regardless of strategy [Ringdahl et al., 2017]. However, in greenhouse conditions one would expect lower success rate and cycle times. This chapter aims to analyze the effect of the approach direction on performance. Additionally, a failure analysis methodology of the different approach strategies provides several conclusions for implementation in robotic harvesters.

Previous research on human-robot collaboration for target recognition has indicated that improved performance can be achieved by including a human in the interpretation of the visual information received from the robot sensors [Berenstein & Edan, 2012; Bechar & Edan, 2003]. Therefore, another aim with this work is to measure the ability of human observers to provide insights on approach directions based on overview images taken in the field.

7.2 Approach strategies

An approach strategy is the method of finding an approach waypoint from which visual servoing will not lose the fruit due to occlusion. It is computed using the approach direction θ_1' and distance to fruit d (Figure 7.1). Two approach strategies are proposed and evaluated:

- **Single approach strategy**: the robot attempts only one approach direction which is considered to be the best one with least occlusion. This direction can be obtained either by advanced algorithms that map the environment or by a human operator doing an

[1] http://www.sweeper-robot.eu

assessment in-field or remotely by looking at images taken by the robot at the overview waypoint.

- **Multiple approach strategy**: the robot autonomously attempts each approach direction from a sequence of a-priori defined approach directions until it finds one that leads to a successfully reached fruit.

This section describes how approach waypoints are calculated, and provides a detailed description of the two strategies.

7.2.1 Approach waypoint calculation

To calculate the pose of the approach waypoint, the following information must be known: the location of the pepper, the approach direction, and the visual servoing distance, i.e. the distance to the fruit from the approach waypoint (marked as d in Figure 7.1).
Given position $P_i(x_i, y_i, z_i)$ of (the surface of) fruit i at approach direction $\theta_i = 90°$ the position for approach waypoint $W_i(x'_i, y'_i, z'_i)$, located at a predefined distance d from fruit i at an approach direction θ'_i, is calculated according to:

$$x'_i = x_i - (r+d) * cos(\theta'_i) \tag{7.1}$$

$$y'_i = y_i + r - (r+d) * sin(\theta'_i) \tag{7.2}$$

$$z'_i = z_i \tag{7.3}$$

where r is the fruit radius, and d, θ'_i, as described in Figure 7.1.

7.2.2 Proposed approach strategies description

Two approach strategies are evaluated, based on the conclusion from the previously published laboratory experiment ([**ringdahl2017strategies**] attached in Appendix 10.2). In the **single approach strategy**, a single approach direction θ'_i should be determined for each fruit i. If possible, approaching from front ($\theta'_i = 90°$) is preferred [Bac et al., 2016], otherwise the least occluded direction should be chosen (more details in Section 7.3.2). The approach cycle starts with moving the end-effector to a pre-defined overview waypoint $W_0(x, y, z)$ where the location of all visible fruits are recorded. From there, using the position and selected approach direction of each target fruit i, the approach waypoint $W_i(x'_i, y'_i, z'_i)$ is calculated according to Equations 7.1-7.3. The control unit plans a path for the end-effector to the first waypoint. After reaching it, a visual servo procedure guides the manipulator towards the target until the end-effector reaches the fruit. If the manipulator is able to reach the target fruit, the fruit is marked as reachable. If the manipulator cannot reach the fruit for some reason, e.g. the view of the fruit is lost during visual servoing or the controller is not able to plan a path there, the fruit is marked as unreachable. After the fruit has been marked as either reachable or unreachable, the approach cycle for fruit i ends. The next approach cycle starts with the robot moving to the approach waypoint of the next fruit, W_{i+1}. The cycle ends when all fruits detected from the overview image have been attempted to be approached. The left part of Figure 7.2 shows a flowchart of this approach strategy.

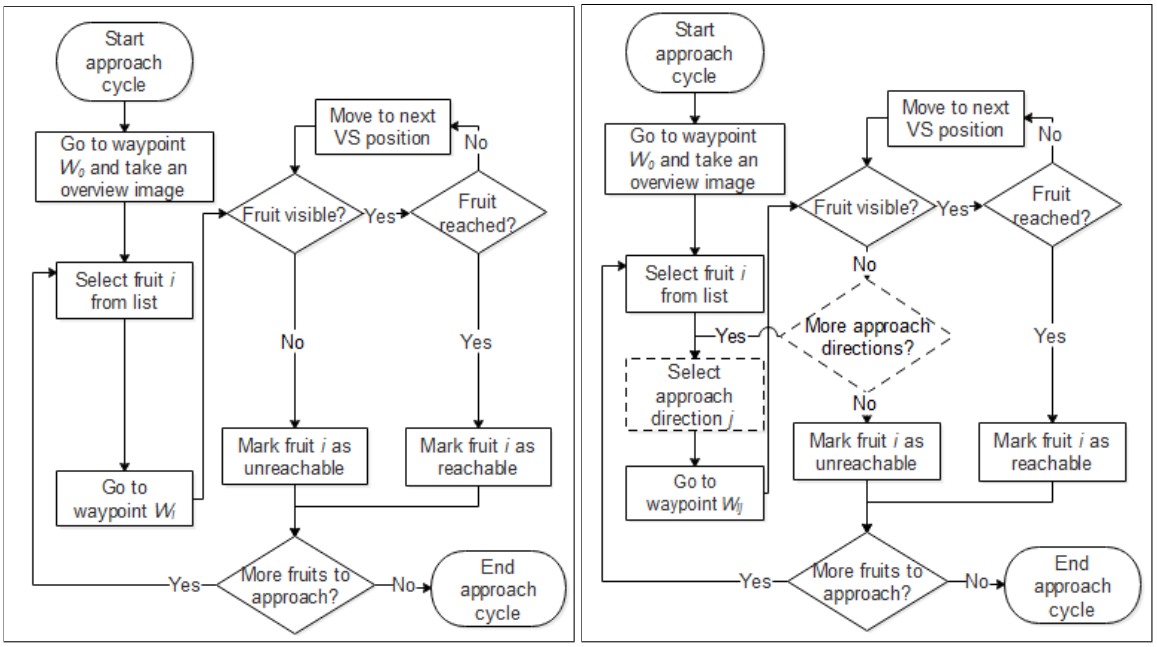


Figure 7.2: Flowchart describing the two different approach strategies. Left: single approach direction; Right: multiple approach direction (differences marked with dashed lines).

In the **multiple approach strategy**, the best approach direction θ'_i is unknown, and therefore must be searched from a list of predefined potential approach directions $\theta'_{i,1} \dots \theta'_{i,k}$ as seen in Figure 7.3. For each target fruit i and potential approach direction θ'_{ij} the control unit calculates the path of the robotic manipulator to a waypoint $W_{ij}(x'_{ij}, y'_{ij}, z'_{ij})$ according to Equations 7.1-7.3. The manipulator moves to each waypoint in turn for the first fruit until the fruit is marked as reachable or all waypoints have been tried. In the case that all approach directions $\theta'_{i,1} \dots \theta'_{i,k}$ have been attempted without being able to reach the fruit, the target fruit is marked as unreachable. After success or fail, the path to the waypoint $W_{i+1,1}$ for the next fruit and its first approach direction $\theta'_{i+1,1}$ is calculated. The right part of Figure 7.2 shows a flowchart of this approach strategy.

7.3 Experimental methods

Two experiments were conducted to compare the performance of the approach strategies. In the first experiment, the multiple approach strategy is compared to the single approach strategy assessed by a human operator placed in the field. In-field human assessment was selected as opposed to fully autonomous algorithms due to lack of sensor technology and algorithms to date. Current algorithms are often partially based on manual data [Barth et al., 2016] and not sufficiently accurate or fast to map greenhouse environments for determining possible approach directions. Stems are not detected in the system used in this work and are thereby not avoided unless they block the view of the fruit.

The second experiment measures the ability of human observers to provide insights to approach directions based on overview images collected in laboratorial and greenhouse conditions.

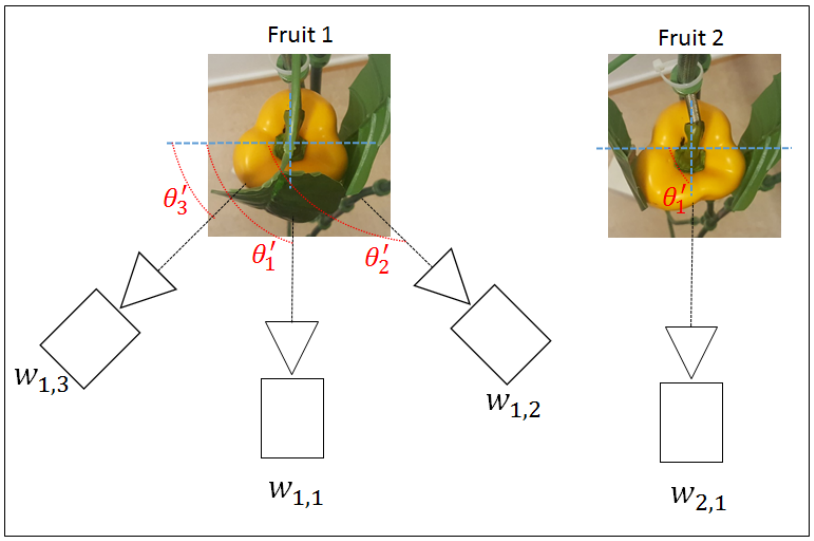


Figure 7.3: For each target fruit i and potential approach direction θ_j' the control unit calculates the path of the robotic manipulator to a waypoint $W_{ij}(x_{ij}', y_{ij}', z_{ij}')$ in the multiple approach strategy. The figure illustrates a center-right-left approach.

7.3.1 Equipment

A 6DOF robotic manipulator Fanuc LR Mate 200iD equipped with an eye-in hand iDS UI-5250RE RGB camera and a Sick DT20HI displacement measurement laser sensor was placed in-front of each scene. The end-effector used in the greenhouse experiments was slightly different than the one used in the previous lab experiment [Ringdahl et al., 2017]. Primarily the lab version had a suction cup mounted in front that could touch the peppers while the greenhouse version lacked this. Therefore, the greenhouse version never touched the peppers but stopped just before it reached the pepper. In laboratorial conditions, an artificial plastic pepper crop with yellow plastic fruits and green plastic leaves were used for the experiments. An example from the laboratorial and greenhouse setups can be seen in Figure 7.4.

The workflow of the robot was implemented using a generic ROS software framework for development of agricultural and forestry robots previously developed [Hellström & Ringdahl, 2013]. The framework is constructed with a hybrid robot architecture, using a state machine implementing a flowchart as described by Ringdahl et al. [Ringdahl et al., 2016].

7.3.2 In-field experimental protocol

The experimental protocol used for the robotic experiments in the greenhouse resembles the previously described laboratory experiment protocol [Ringdahl et al., 2017], while addressing some of the conclusions and their implementation in greenhouse conditions. Table 7.1 outlines the protocol used for laboratorial and greenhouse experiments respectively. The difference between greenhouse and laboratorial conditions can be seen in Figure 7.5, which shows overview images taken from the respective environment.

The greenhouse experiment was conducted in a research greenhouse at Sint-Katelijne-Waver, Belgium, early in the season. The robotic manipulator and its sensory system were placed in an aisle and manually centerd in front of a pepper or a cluster of peppers, defined as a

Figure 7.4: The experimental setup in the greenhouse (left) compared to the previous experiment in laboratorial conditions using artificial fruits and leafs (right).

scene. First, the locations of all the fruits in the scene were registered. In this experiment the system was not equipped with a RGB-D camera, which would typically be used for automatic registration of fruits locations. Instead, the manipulator was manually positioned in front of each fruit before the experiment began. The end-effector was brought close to the fruit and the position was determined by adding the distance to the fruit, given by the laser mounted on the end-effector (Figure 7.6), to the current end-effector position. The system used in the experiment lacked a sensor to accurately measure the fruit radius, but even with such sensor it can be difficult to estimate the correct radius due to occlusion of parts of the fruit. Peppers in a greenhouse have a natural variation in size but normally have an average radius $r = 0.04m$ [Bac et al., 2014b; Tadesse et al., 2002], so this value was used for the calculations in Equations 7.1-7.3. All fruit locations were added into the system in form of spheres so that the robot could avoid them when planning a path. Each fruit was evaluated for it's approachable direction θ_i' by a human observer who examined the fruit and chose the front direction $\theta_i' = 90°$ if it was clear and if not one of the angular approach directions. Next, an approach cycle was performed five times for each fruit; once for the single approach strategy and four times for the multiple approach strategy, one for each potential approach direction (Table 7.1). Since ripe peppers are in high color contrast with the foliage and branches around them, color is the most useful visual cue. Therefore the visual servoing used for each approach attempt employs a color based blob detection algorithm for the continuous detection of the target. Minimum and maximum thresholds of the blobs' bounding and inner circles radii were introduced to limit false positive artefacts. Different visual servoing distances were evaluated in this experiment as opposed to a fixed distance in the lab, as well as a different number of approach directions and their order. The list is ordered in two possible ways: center first (CF) or left first (LF) and can have either 3 or 5 directions. Previous results [Ringdahl et al., 2017] showed no significant difference for different robotic maximum speeds, therefore each configuration is performed at 25% of maximum speed to assure safe operation in the field.

Measurements

At the end of each approach attempt, the following measures were registered: the result of the attempt (success/failure) and the reason of any attempt failure (collision, planning failure, or lost fruit from sight during visual servoing). For each fruit, additional measures were

Table 7.1: Experimental protocol in the greenhouse compared to the laboratory experiments presented in [Ringdahl et al., 2017].

	Laboratorial experiment	Greenhouse experiment
Single approach strategy	Best approach direction from the set {45°, 90°, 135°}	Best approach direction from the set {45°, 90°, 135°}
Multiple approach strategy	Potential approach directions: – **LF3:** $\theta_j = \{45°, 90°, 135°\}$ (left-center-right) – **CF3**: $\theta_j = \{90°, 45°, 135°\}$ (center-left-right)	Potential approach directions: – **LF3:** $\theta_j = \{45°, 90°, 135°\}$ (left-center-right) – **CF3**: $\theta_j = \{90°, 45°, 135°\}$ (center-left-right) – **LF5**: $\theta_j = \{25°, 57.5°, 90° 122.5°, 155°\}$ (center-left-right) – **LF5**: $\theta_j = \{90° 25°, 57.5°, 122.5°, 155°\}$ (center-left-right)
Maximal Robotic speed	50% and 100%	25%
Visual servoing distance	5 cm	9 cm/15 cm
Measures per approach attempt	– Success/failure	– Success/failure – Reason for failure
Measures per fruit	– Fruit approach cycle time – Number of attempted approach directions	– Fruit approach cycle time – Reachable/unreachable – Number of attempted approach directions

registered: the fruit approach cycle time and the number of attempted approach directions. Fruit approach cycle time is the time it takes from when the robot starts moving to the first waypoint of a fruit until it has been marked as reachable or unreachable. Additionally, the following measures were calculated for each approach strategy and are presented in form of descriptive statistics in the results section:

- **Cycle time relative increase**. Defined as the ratio between the average fruit approach cycle time for the multiple approach strategy and the single approach strategy. It measures the impact on the fruit approach cycle time given an approach direction known a-priori.
- **Ratio of reachable fruits** in each approach strategy. Allows a comparison between the success ratio of each approach strategy.
- **Ratio of successfully approach attempts**. Allows insights into which approach directions are most successful.
- **Approach attempt failure ratio**. Provides insights into the reasons an approach attempt fails and the frequency at which failures occur.

Additionally, the statistical significance of the differences in the value of the measures was calculated. The fruit approach cycle time is analysed in a box-cox transformed linear regression [Sakia, 1992]. The reachable fruit rate as function of approach strategy and visual servoing distance is analysed in a logistic regression [Hosmer Jr et al., 2013].

Failure analysis

To be able to investigate the reasons leading to unsuccessful approaches, the following failure analysis methodology is presented. The failure analysis looks into the reasons why an approach strategy fails, resulting in a fruit being marked as unreachable, as well as the reasons for failure of individual approach attempts in the multiple approach strategy. The main failure reason this chapter addresses is related to occlusion, but one should consider

Figure 7.5: An overview image taken from the robot's camera looking at a laboratorial scene (left) and a greenhouse scene (right).

Figure 7.6: The position of each pepper was measured by manually moving the robotic arm close to the fruit and using the laser to estimate the distance to the fruit (the end-effector used in the greenhouse lacked the suction cup seen in the picture).

the tight space between plant lanes in which the robot has to operate in a safe manner, minimizing the harm caused to the plants around. Therefore, other failures, caused by the planner, are considered and measured. As mentioned in Section 7.3.2 the visual servoing failures are registered at the end of each unsuccessful approach attempt. The failure reasons are then separated into three categories:

- **Occlusion related failure**. Cases where the fruit was lost from sight during the visual servoing stage.
- **Visual servoing planning failure**. Cases where the next coordinate generated in the visual servoing path planner cannot be reached due to physical constraints of the manipulator, collision between the robot and that the environment, or the planner fails to find a solution.
- **Approach waypoint planning failure**. Cases where approach waypoint W_{ij} is unreachable due to the physical constraints, collision between the robot and the environment, or that the planner fails to find a solution.

Looking into the different reasons causing an approach to fail is important from two aspects. First, it provides insights into the reasons for failure due to non-occlusion related issues. It also provides an estimation of the expected success rate of an approach attempt that has been made with no prior knowledge about the best approach direction as in the multiple-approach strategy. The failure analysis presents the approach attempt failure ratio divided into the

three categories listed above. It also provides analyses of the difference in success rates for different approach directions with and without planning related failures.

7.3.3 Overview images evaluation protocol

To evaluate the human ability to provide insights into possible approach directions θ'_i or limiting the number of directions that should be attempted by determining directions that are not approachable, two datasets containing images taken at the overview waypoint W_0 were collected:

1. **DB-LAB** - Laboratorial images acquired at the overview waypoint W_0 in the laboratorial experiment described in our previous publication [Ringdahl et al., 2017].
2. **DB-Greenhouse** - Greenhouse images acquired mid-season in a commercial greenhouse in Ijsselmuiden, Netherlands using the same camera as in DB-LAB mounted on a Fanuc LR Mate 200iD/7L manipulator. The images were automatically acquired from 14 waypoints [Kurtser et al., 2016].

Images from the greenhouse experiment described in Section 7.3.2 were not used for the evaluation due to the excess number of non-visual servoing related failures.
A questionnaire was handed out to 13 human observer participants and included images from the two datasets, divided into two parts:

1. **Laboratory part**: contained 6 images: one example image (Figure 7.7) and five from DB-Lab (one of them is presented in Figure 7.5).
2. **Greenhouse part**: contained 6 images from DB-Greenhouse. The images were randomly selected with the following limitations: a) image was taken in daytime; b) at least 2 ripe peppers were visible from front view. One of the images is presented in Figure 7.8.

The fruits in the images displayed to the participants were bounded by red boxes. The participants were asked to mark all directions (if any) from which, in their opinion, each pepper could be successfully approached (left/front/right). According to the instructions a pepper must be visible in the robot camera for a successful approach. They were given an example image (Figure 7.7) with views of one fruit from the laboratorial setting from the three potential approach directions. Even if a fruit is partly occluded by a stem it should still be regarded as approachable, as seen in Figure 7.7 (right). By purpose the participants were not given a definition of how visible a fruit must be to be approached. The participants were also asked to indicate if they had any previous experience working in greenhouse environment or doing research within agricultural robotics.
By processing the answers an *observed approachability list* AP_{ijk} which defines for each fruit i marked in the overview image if it is reachable from the approach direction θ'_j, by the vote of observer k is generated. Given the results of the approach attempt performed in the laboratory a *ground truth approachability list* AP^*_{ij} is defined as the rounded ratio of successful attempts for fruit i from approach direction θ_j, to the overall number of attempts performed for that fruit. Given the ground truth approachability and the *observed* approachability list AP_{ijk}, the contribution of a human participant's ability, from an overview image only, to predict

Figure 7.7: Example lab image of a fruit from three different viewpoints given to the questionnaire participants. They were told that the left and right viewpoints are reachable, while the center one is not.

if an approach direction will lead to a successful approach is evaluated. The evaluation is performed using precision and recall measures, defined as follows:

$$Precision = N_{TP}/(N_{TP} + N_{FP}) \tag{7.4}$$

$$Recall = N_{TP}/(N_{TP} + N_{FN}) \tag{7.5}$$

Where:

- N_{TP} is the number of correctly classified approachable directions as approachable by a participant.
- N_{FP} is the number of incorrectly classified unapproachable directions as approachable by a participant.
- N_{FN} is the number of incorrectly classified unapproachable directions by a participant.

Due to the lack of a ground truth approachability list AP^*_{ij} in DB-Greenhouse, precision and recall are not calculated. Instead the reliability of agreement Fleiss' Kappa [Fleiss & Cohen, 1973] is measured for both greenhouse and laboratorial conditions. This allows us to gain initial insights into the inter-rater agreement between participants on the choice of approachable directions. Various scales of Fleiss' Kappa are accepted in the literature. In this chapter we follow Landis and Koch's guidelines [Landis & Koch, 1977].

7.4 Results

7.4.1 Approach strategies cycle time and success rate comparison

A single plant row was measured orientated approximately southwest – northeast. The platform was placed in both aisles on the both sides of the row in the aim to cancel the influence of the foliage growth direction (due to sunlight direction) on approachability. A total of 18 peppers facing the first aisle (approximately northwest), divided in five scenes, were approached from 9 cm and a total of 12 peppers facing the second aisle (approximately southeast), divided in 3 scenes, were approached from 15 cm. Each pepper was attempted to be approached five times, one for each approach strategy described in Section 7.3.2. This

Figure 7.8: An overview image from a greenhouse scene. The questionnaire participants were asked to determine all possible approach directions for the peppers marked with red boxes.

creates a dataset with 150 fruit approaches (30 fruits multiplied by 5 approaching strategies combinations) in eight scenes with 2-6 peppers in each scene. In total, 244 approach attempts were made. The total ratio of reachable fruits was 86%. 84% and 87% of the fruit are reachable in the single approach strategy and in the multiple approach strategy respectively (Table 7.2). As seen in Figure 7.9, the most approachable direction θ_i was $90°$, with 74% successful approaches made from that direction. This number is as expected lower the the fruit approach success since a single fruit was approached more then once till success. Very few approaches were made at 135 and 155 degrees, implying that the fruit was successfully approached at one of the other approach directions (right was always tested last in all approach strategies). Removing all cases where the robot failed to plan a safe path or collided with another fruit in a cluster during approach shows that $125°$ is the most approachable direction (although only approached 7 times) with $90°$ the second best (approached 93 times). Comparing performance of the single approach strategy to the multiple approach strategy revealed that the fruit approach cycle time increased when a multiple approach strategy was applied by 8% for reachable and by 116% for unreachable fruits respectively (Table 7.2).
In 16% of the cases for the single approach strategy, the fruit was unreachable even though it was assessed manually in-field as the best approach direction (see Section 7.4.1 for an analysis of failure cases). For the single approach strategy, the fruit approach cycle time of unreachable fruits was 33% shorter than for the reachable fruits. In the multiple approach strategy, the fruits were unreachable in 13% of the cases and the fruit approach cycle time was 33% longer for unreachable fruits than for reachable fruits.
Table 7.3 shows the average fruit approach cycle time in seconds for the different approach strategies described in Table 7.1 for the two visual servoing distances (9 and 15 cm). The single approach strategy yielded the shortest overall time for 9 cm visual servoing distance with 13% shorter time than the next best strategy, CF3. At 15 cm approach distance, the CF3 and single approach strategy took approximately the same amount of time on average. As can be seen in Figure 7.10, CF3 and CF5 needed on average only one approach to succeed, meaning that most of the time it only needed to go to the center position without needing to

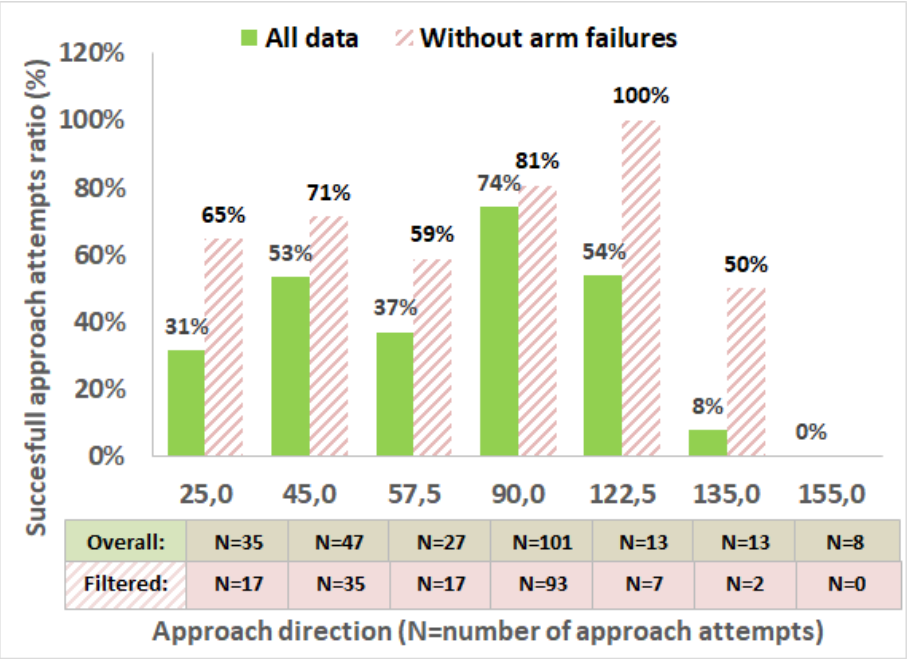


Figure 7.9: Rate of successfully approach attempts by approach direction θ_i', and the number of times each approach direction was attempted. In the right bars, cases where the robot failed to plan or collided with other fruits in the scene were removed.

Table 7.2: Average fruit approach cycle times in seconds for the different approach strategies.

Approach strategy	Reachable	Unreachable
Overall	16.02 (N=129, SD=7.67)	18.85 (N=21, SD=12.02)
Single	15.04 (N=25, SD=5.74)	9.99 (N=5, SD=11.15)
Multiple	16.26 (N=104, SD=8.08)	21.61 (N=16, SD=11.18)

try another direction. The LF5 method needed most trials, with an average of two approaches before success.

Figure 7.11 shows the success rate for the different approach strategies (Table 7.1) for the two visual servoing distances in the greenhouse. In general, the success rate was between 75-95% with slightly lower success when approaching from 15 cm than when approaching from 9 cm, with the exception of the LF5 strategy. Using five approach directions yielded slightly higher success rate than when using three approach directions.

Failure analysis of greenhouse data

Of the total 244 approach attempts, 30 attempts were part of a single approach strategy and 214 were part of the multiple approach strategy. Forty-seven percent (115 approaches) of the approach attempts failed, indicating the importance of determining a correct approach. The most common reason (41%) for failure was due to the robot planner not being able to find a safe path to the fruit during visual servoing or while moving to the next approach waypoint. 37% of the failed approach attempts were because the fruit was lost from view along the visual servoing. The remaining 22% of the failures were due to a collision between the robot and another fruit in the scene. Figure 7.12 presents the ratio of all approach

Table 7.3: Average fruit approach cycle time in seconds for the different approach strategies described in Table 7.1 for two different visual servoing distances.

		Success		Failure		Overall	
	Potential approach direction	9cm	15cm	9cm	15cm	9cm	15cm
Multiple approach Strategy	CF3	13.2 (SD=4.5)	13.5 (SD=3.7)	24.1 (SD=29.8)	25.3 (SD=9.3)	14.4 (SD=9.1)	16.4 (SD=7.4)
	LF3	14.0 (SD=4.5)	19.7 (SD=12.5)	23.4 (SD=8.6)	20.4 (SD=13.9)	15.6 (SD=6.2)	19.8 (SD=12.1)
	CF5	15.8 (SD=8.9)	19.5 (SD=11.4)	9.5 (SD=0)	23.5 (SD=13.3)	15.5 (SD=8.7)	20.2 (SD=11.2)
	LF5	15.7 (SD=6.7)	21.5 (SD=7.9)	19.4 (SD=4.6)	15.4 (SD=0)	16.1 (SD=6.5)	20.9 (SD=7.8)
Single approach Strategy	-	13.4 (SD=4.7)	17.9 (SD=6.4)	8.2 (SD=11.1)	11.2 (SD=13.4)	12.8 (SD=5.5)	16.3 (SD=8.5)

attempts outcomes for the two approach strategies separately. Approach attempts made from an approach direction obtained by manual in-field assessment as part of the single approach strategy was successful in 84% of the cases. In 10% of the unsuccessful approaches the fruit was lost during visual servoing (i.e. the fruit was no longer detected) while the remaining cases (6%) were a result of a collision between some part of the robot (usually the end-effector) and another fruit in the cluster or that the planner could not find a safe path while moving to the next approach waypoint.

Looking at each individual approach attempt as part of the multiple approach strategy, only 49% of the approaches were successful, while 18% were lost in visual servoing and 10% were lost as a result of planning failure during visual servoing. While moving to the next approach waypoint, 23% of all approach attempts failed due to planning failure or collision.

Figure 7.9 presents the approach success rate by approach direction θ'_i. Removing all the cases where the robot failed to plan a safe path or collided with another fruit in the cluster during approach shows that almost no planning related failures happened at $90°$ (81% compared to 74%) while the side views caused a significant amount of planning failures, increasing as function of $|\theta'_i - 90|$ (e.g., 65% compared to 31% for $25°$). This supports the assumption made in Section 7.2.2 that approaching from front $90°$ should be preferred as long as this approach is not occluded. The low rate of planning related failures for $90°$ can be explained by the fact that this scene contains much fewer fruits and other obstacles from the front approach direction, making the planning task less complex and the risk of colliding with obstacles lower. When side approach directions are applied ($|\theta'_i - 90| > 0$) the robotic manipulator needs to reach deeper into the plant making it more likely to cause collisions between the robot and other fruits. Additionally, it seems the approach directions from the right ($\theta'_i > 90°$) fail more often due to planning than from the left. Important to note that these differences are not due to any botanical changes since these are planning failures, which are not connected to the growth of foliage in greenhouse conditions. Since much fewer attempts were made from the right this needs to be verified in future work.

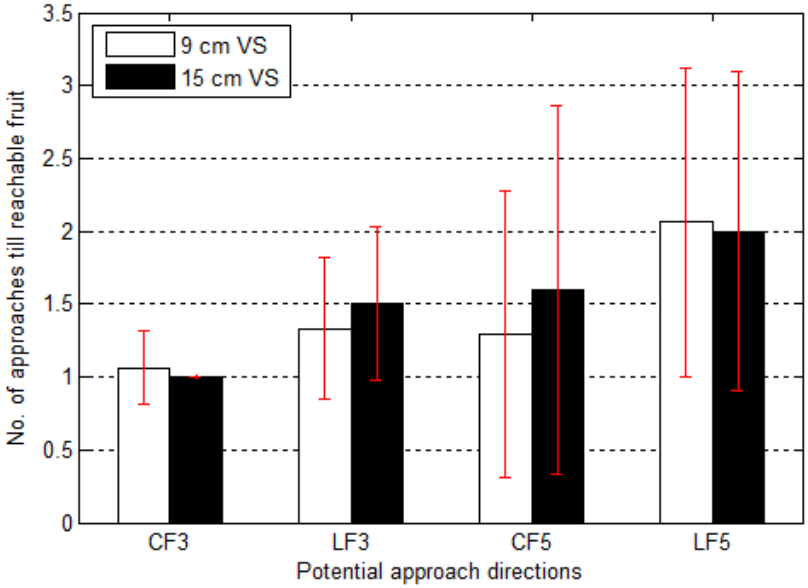


Figure 7.10: Average number of approaches till successful fruit approach for the different potential approach directions of the multiple approach strategy described in Table 7.1 for two visual servoing distances.

Approach cycle times and fruit reachability analyses

The results of a Box-Cox transformed ($\lambda = 0.44$) fruit approach cycle time regression shows statistical significant ($sig < 0.0001$) differences between the multiple approach strategies (CF and LF) and the single approach strategy. The multiple approach strategy adds on average 3.5 seconds to the fruit approach cycle time as compared to the single approach strategy. No statistical significant differences were found between the CF and LF approach direction sorting ($sig = 0.72$). The number of approach directions were not statistically significant either ($sig = 0.18$). The visual servoing distance was found to be a significant factor ($sig = 0.001$), increasing the fruit approach cycle time by 0.28 seconds for each extra cm of visual servoing. The approach cycle time of a reachable fruit was on average 2.4 seconds longer than an unreachable fruit. Significant differences in approach cycle time between scenes were found,

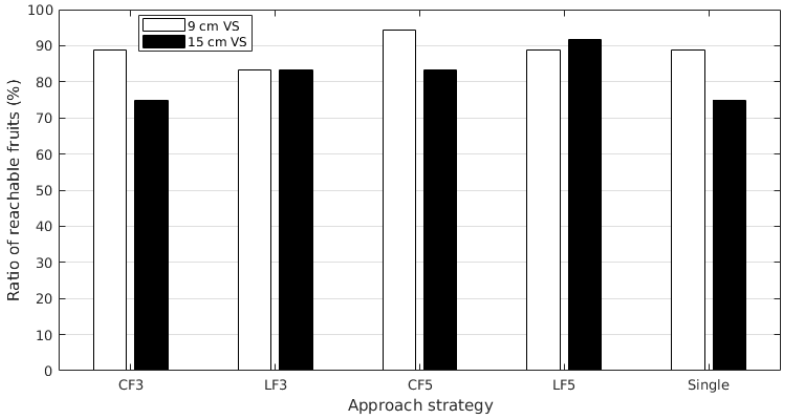


Figure 7.11: Success rate for the different approach strategies described in Table 7.1 for two different approach distances.

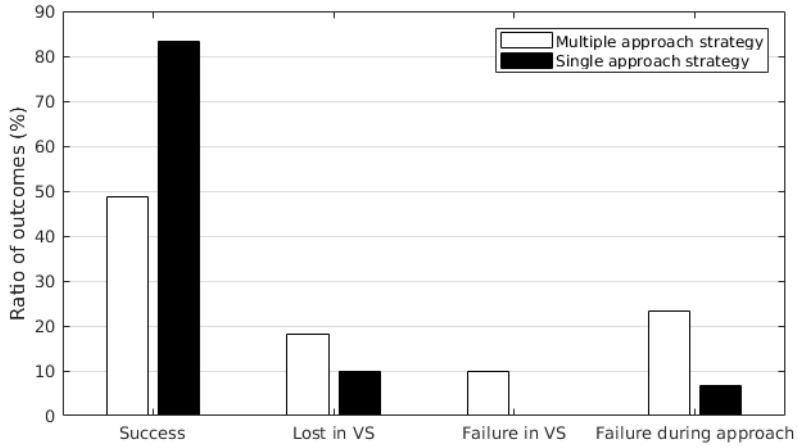


Figure 7.12: Ratio of all approach attempts outcomes for the two approach strategies separately. Lost in VS means the fruit was no longer detected while doing visual servoing. Failure occurred due to not finding a safe path or due to collision between the end-effector and a fruit.

implying that additional variations between the scenes also affects the cycle time. This requires further investigation on larger datasets.

Looking into the interaction between the approach strategy and if the pepper was reachable or unreachable (Figure 7.13), reachable fruits yield shorter approach cycle times than unreachable fruits for the multiple approach strategy. In the single approach strategy, unreachable fruits yield shorter approach cycle time than reachable fruits. These results support the conclusions drawn from Table 7.2 and found statistically significant ($sig = 0.001$).

The result of a logistic regression on reachable fruit rate as function of approach strategy and visual servoing distance revealed no statistically significant differences. This might be due to the relatively small dataset gathered in the greenhouse experiment.

7.4.2 Human contribution to approach direction assessment

In both the greenhouse and the previous laboratory experiments, the average fruit approach cycle time increased for the multiple approach strategy as compared to the single approach strategy using in-field assessment by human observer. In this section the human ability to determine the best approach direction from overview images is evaluated.

All 13 participants answered all questions. Four participants had prior experience of work in greenhouse conditions and 9 never worked in greenhouses before.

In laboratorial images collected, the average precision per person (Equation 7.4) was 92% (min=85%, max=97%) and the average recall per person (Equation 7.5) was 71% (min=40%, max=83%). No significant differences were found between the four participants who had prior work experience in greenhouse conditions, to the nine who were not.

As aforementioned, the precision and recall on the greenhouse data cannot be calculated due to the lack of ground truth information. Instead, the evaluation includes investigation of the degree of agreeability between the participants, e.g. if all answered front as an approachable

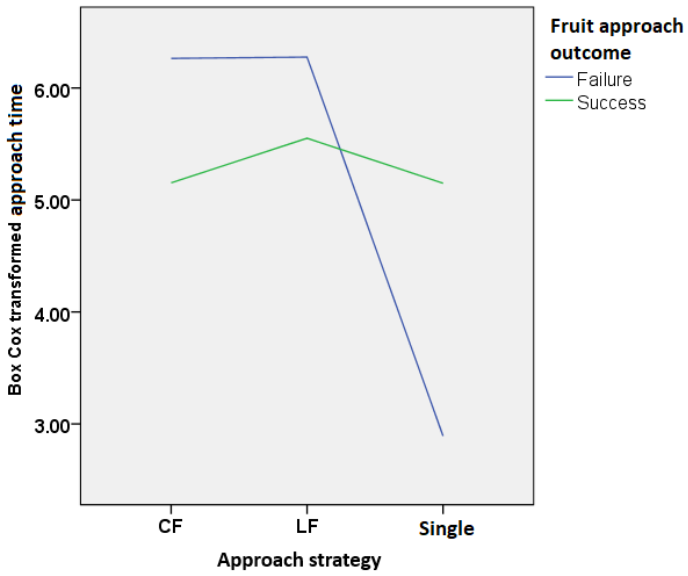


Figure 7.13: Interaction plot of fruit approach cycle time as function of approach strategy for reachable and unreachable fruits.

Table 7.4: Fleiss' Kappa values for data collected in the laboratory and greenhouse.

	Laboratorial conditions	Greenhouse conditions
All data	0.514	0.217
Front approachability	0.739	0.168
Side approachability (left/right)	0.402	0.201
Expert participants all data	0.48	0.37
Non-expert participants all data	0.52	0.15

direction for fruit i. Fleiss' Kappa for the laboratorial images is 0.514, corresponding to moderate agreement level, and for the greenhouse collected data was 0.217 which corresponds to a slight-to-fair agreement. This is a significant reduction in the inter-rater agreement between participants which indicates that as expected the greenhouse task is by far more difficult to analyse. A closer look into the measured Kappa as a function of approach direction and the participants work experience in greenhouse conditions can be found in Table 4.

7.5 Discussion

7.5.1 Approach strategies experiments

The results from the greenhouse experiments support previously reported results [Ringdahl et al., 2017] indicating an increased fruit approach cycle time when using the multiple approach strategy as compared to a single approach direction yielded from manual-infield assessment of human observers. Therefore, additional information about approachable directions is necessary to shorten robotic harvesting times. The data from the greenhouse experiment showed a less prominent time increase (ca 8%) compared to the previously

published laboratorial experiment (ca 40-45%). Even so, it was found to be statistically significant. Longer visual servoing distances was also found to increase fruit approach cycle time. On the other hand, the number of approach directions (3 or 5) or their order in the multiple approach strategy did not significantly influence the fruit approach cycle time. This validates the conclusions reported from the laboratorial experiments.

While the robot eventually managed to approach all peppers in the laboratorial experiment, in the greenhouse each approach strategy resulted in only between 75-95% approachable peppers. The most successful approach direction was from front, supporting the decision to prioritise it in the single approach strategy. Failure analysis showed multiple reasons for failure during an approach attempt. The most common failure (41%) was the inability to plan a safe path in the tight space between plant lanes and stems in the greenhouse. Another reason for failure was that the end-effector collided with another fruit in a cluster while moving towards the next approach viewpoint. This failure is expected to happen much less often in a real application where the peppers are harvested and thereby would not be in the way for the robot when moving towards the next fruit. Only 37% of all failures were due to fruit loss from view due to occlusion during visual servoing towards the fruit. In ideal conditions, one would expect this to be the only reason for not being able to approach the pepper. This is also the only failure one could expect a human observer to predict in field or by observing overview images. If we remove all other reasons for failure (73 cases of the total 244 approach attempts), the robot not reaching the fruit due to occlusion only occurs in 25% of all approach attempts.

Results showed slightly lower reachable fruit rate for the single approach strategy than for the multiple approach strategy. However, no statistically significant difference could be seen in the greenhouse data, probably due to the limited size of the dataset. Since only one approach attempt is made per fruit in single approach strategy, slightly lower fruit reachability is to be expected since it cannot attempt any other approach directions if the first was unsuccessful. Furthermore, the non-occlusion related failures mentioned above were not considered when best approach direction was chosen by an in-field human observer for the single approach strategy.

7.5.2 Human contribution

The evaluation of the human ability to assist in determining the best approach direction in laboratorial conditions revealed high precision (92%), implying that the participants rarely predicted non approachable directions to be approachable. Consequently the participants showed a high ability to detect unapproachable directions. On the other hand, the 70% recall suggests that they tend to miss possible approach directions. The implications of these conclusions are that one should consider humans for reducing the list of possible approach directions used in the search pattern, but should not rely on a human operator to define the best approach direction. The greenhouse experiment showed that an increased number of approach directions does not significantly increase fruit approach cycle time. The implication of this is that the system would be best served by letting humans determine which peppers are unreachable and thereby saving a considerable amount of time, especially since the multiple

approach strategy takes significantly longer time to fail.
A deeper view into the values of the kappa for front/side approach directions reveals significant agreement on the front direction in laboratorial conditions, and lower agreement on the side views. This implies that a human finds the side view to be more complex for approachability prediction. This difference was not found in greenhouse conditions, where the agreement level remained low for both front and side approach directions. In greenhouse conditions though, participants with prior experience in agricultural robotics or greenhouse environments showed greater agreement among themselves compared to the non-experienced participants. No such difference is seen in the laboratorial setting. The results imply that though no significant difference was found in precision and recall in laboratorial conditions for experienced and non-experienced participants, these differences might become significant when the experiment is performed in the more complicated greenhouse conditions where experience can become advantageous.

7.5.3 Limitations

This work aimed to evaluate the human's ability to determine that a pepper is not approachable at all by a robot. As concluded in Section 7.5.2 the laboratorial tests showed that the humans are quite good at this. The data from the greenhouse experiment described in Section 7.3.2 was unfortunately not sufficient to generalize these results to greenhouse conditions due to multiple non-occlusion failures as noted in Section 7.5.1. Future work with a more optimal harvesting system should reduce these failures and allow for a fair comparison.
In this work, stems were not avoided unless they covered significant parts of the fruit and thereby causing the visual servoing to lose the fruit from sight. Future work on selecting approach directions should incorporate stem detection in the robotic harvesting cycle to avoid stems that are between the fruit and the approach waypoint in order not to damage them. Some initial work on detecting and avoiding stems has been done [Bac et al., 2014a] and could be extended to be used in this application in the future.
Sunlight is a significant factor in the botanical growth models of the leaves. As mentioned in Section 7.4.1, the protocol aimed at creating a balanced dataset to cancel the influences of growth direction of the foliage due to different sunlight direction. The results showed that the vast majority of the successful approaches were for $\theta_i' <= 90$, regardless of side of aisle (and thereby sunlight direction). One would expect that if the sunlight direction was a factor in approachability, there would be a difference in which angles are more successful for different sides of the aisle. The results showed no significant differences in which directions were successful and not between the different sides of the aisle, indicating that the sunlight direction did not matter. However, since the visual servoing distance were different for the different sides of the aisle it is not possible to prove explicitly that sunlight did not play a part. Current detection rates (up to 85% [Bontsema et al., 2015; Bac et al., 2014b; Kapach et al., 2012]) allow evaluation of the approach direction strategies independent of detection performance. Any effect of sunlight direction (or other environmental conditions) on the visual servoing performance is not addressed in this chapter (e.g. if the fruit was lost from sight due to the sunlight).

7.6 Conclusions

The main contribution of this work is to suggest and compare approach strategies, while measuring cycle times and success rate as well as analyzing causes of approach attempt failures. For the proposed greenhouse environment and hardware configuration it was shown that a multiple approach strategy results in 8% longer cycle time than a single approach strategy and that the most common failure was the inability to plan a safe path in the tight space between plant lanes and stems in the greenhouse. The exact measures might vary between different harvesting systems, but the evaluation method of different approach strategies can be applied to any robotic system using visual servoing in highly cluttered environments.

The issue of human in the loop is a question that is often raised in the literature as a possible alternative to fully autonomous systems which have limited performance. The slight increase in fruit approach cycle time in the multiple approach direction strategy is highly unlikely to justify the inclusion of a human observer in the field to assess the best approach direction. Therefore, it is important to develop algorithms that will be able to provide this information in shorter times and lower costs. For other hardware configurations and greenhouse environments, the increase might be significant enough to justify a human in the loop. To identify these cases one should consider comparing the different approach strategies using the methods proposed in this work.

Since the ability to detect fruits that are unreachable from overview images can significantly shorten the overall harvesting cycle time methods should be explored to determine this. More research is needed to investigate the times it takes for a human to detect unreachable fruits in order to analyse the economics of human robot collaboration for the proposed tasks.

Given the high failure rate of the single approach direction due to non-occlusion related failures, one should consider a hybrid approach strategy. This strategy would start with using the single approach strategy, since it yields the shortest cycle times. If the direction assessed to be the best one fails, the multiple approach strategy can be applied to search for a possible approach direction. The hybrid approach aims to increase the ratio of reachable fruits and should be investigated further. If no information about the best approach direction is available, the multiple approach strategy should be considered to lower the rate of unreachable fruits. If a fixed approach direction is the only applicable strategy then the proposed approach direction for sweet pepper harvesting is from the front ($90°$).

7.7 Research questions answered

The following research questions defined in section 1.2 have been met:

RQ4.1: What is the value of knowing the correct approach direction vs searching for it? The time reduction as a result of knowledge of the correct approach direction has been evaluated in the described robotic setting. A reduction of 40% and 8% in laboratory and greenhouse conditions respectively has been achieved.

RQ4.2: Can humans identify the correct approach direction? The ability of humans to evaluate the approach direction from images has shown limited ability however, human observers are able to detect unapproachable directions.

RQ4.3: What are the common failures occurring during fruit approaching in visual servoing? Main types of failures have been identified and separated into two groups- visual failures and mechanical failures. The frequency of such failures was recorded and analyzed.

Conclusions & Recommendations

8

> *I am turned into a sort of machine for observing facts and grinding out conclusions.*
>
> — **Charles Robert Darwin**
> (English naturalist, geologist and biologist)

This thesis presented several methods for improving performance of autonomous harvesting robots by introducing planning of sensing actions into the robot's operation. The research contributes with developments related to planning the static locations of the sensing points, introducing dynamic locations choice, introducing task planning after the locations have been derived, and planning the sensing for harvesting and approach towards the fruit. All four enhancements have yielded reduction in harvesting cycle time and increase in detectability – both major bottlenecks preventing autonomous harvesting robots from penetrating the market.

8.1 Main results

The statistical analysis of fruit detectability revealed that choosing the right combination of viewpoints leads to a significant increase in detection in complex mid season environment. This combination of viewpoints does not remain static between growing methods, or even within the season and therefore should be analyzed dynamically. A dynamic sensing strategy results in increased detection and decreased costs as compared to the single viewpoint strategy employed today. Optimizing the harvesting sequence resulted in 12% reduction in cycle times. Cycle times can be further reduced by choosing the correct approaching direction.

8.2 Recommendations

It is recommended, in future developments of robotic harvesters to:

- Perform statistical analysis of best viewpoint combinations. Each crop, environment and robotic setting will have a different set of guidelines and therefore it is important to identify the best-fit viewpoints for a given combination of crop-environment-robot.
- Incorporate dynamic sensing. If sensing is limited from a single viewpoint, an algorithm similar to the one presented in this thesis should be included. The algorithm should compare the benefit of additional viewpoints to the cost and accordingly decide if additional viewpoint(s) is needed and where should they be located.
- Perform calculation of best harvesting sequence. Robotic harvesters should sequence the targets and the sensing operations.
- Evaluate approach strategies. In occluded settings, identification of proper approach directions is crucial for visual servoing based fruit approach and should be integrated.

Some limitations that are important to note as well as future enhancements that should be considered are outlined in the following subsections.

8.3 Limitations

In addition to the discussed limitations of each of the developed methods as detailed in Chapters 4 - 7, general assumptions have been made regarding the configuration of the robotic manipulator and its operation as detailed below.

The algorithms developed in this thesis assume an eye-in-hand configuration of the robotic harvester. Even though, as discussed in the introduction and background, this is the most common configuration of a robotic harvester, other static configurations may not allow dynamic sensing points or visual servoing to take place. For these configurations, other methods of dealing with occlusions should be provided. These methods can include (as outlined in Chapter 5 and Chapter 7), mechanical blowers or growing techniques that will limit the amount of occlusion.

Another assumption in this research is real-time operation of the robotic system (assuming negligible processing times). While some previous robots [Bac et al., 2014a] resulted in slow operation times and increased cycle times due to limited computational power (e.g., path planning algorithms, image processing), recent research has shown this issue has been nearly solved. Since computational power is expected to increase this has been disregarded in the calculation of viewpoint processing and task planning. The entire focus in this research is on minimizing arm movements by planning of sensing.

In this research, both the dynamic sensing algorithms and the sequencing of targets employs a heuristic resulted in increased performance. However, the results could be further improved by developing optimal algorithms.

Furthermore, the calculation of viewpoints, task planning, and approaching was limited in this research regarding consideration of safety operations within the greenhouse. Safety can be considered in two aspects – safe operation for the humans working next to the robotic harvester, as well as safety in harvesting without harming vegetation. While all safety measures for humans have been implemented (including operating in lower speeds than maximal speed, emergency button configuration, etc.) very little has been done for obstacle avoidance, e.g., not hurting stems. Several algorithms have been developed in the literature to address this issue [Bac, 2015], but integration of such functionality might affect the utility functions due to the change in path planning between harvesting points. Additionally, this thesis focuses on improvements of autonomous harvesters that assumes no human intervention, The work performed in Chapter 7 on evaluation of a human-robot collaboration strategy is limited and is conducted for comparison to autonomous strategies.

Finally, this research assumes that the harvested fruit is directly transported to a bin (assuming there are no additional movements from the fruit location to a bin). Since additional movements prolong harvesting times this should be the common practice of using solutions such as a plastic tube collecting the harvested peppers to the bin (e.g., [Bac et al., 2017]).

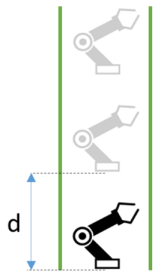


Figure 8.1: Robot location within the lane. d is the distance of advancing along the lane.

Figure 8.2: High occlusion level crop.

8.4 Future work

Ongoing work is related to planning of sensing locations while the robotic harvester platform advances along the lane (Figure 8.1). In this research the robotic harvester platform is assumed to be placed at a specific location along the lane. The exact location has not been addressed. Initial calculations to derive the platform location were made in Chapter 4 but they were not sufficient for deriving a definitive conclusion for questions such as - at what distance (d in Figure 8.1) along the lane should the robotic cart stop in the search for sweet peppers and should this decision be dynamic as well (change according to previously sensed data).

The application of the developed methods to other crops requires adaptions as noted below. The viewpoint analysis can be applied using the presented methodologies to crops with high presence of occlusions (e.g., citrus [Bulanon et al., 2009], grapes (Figure 5.2) and tomatoes (Figure 8.2)). Crops with high inherent visibility (e.g., Cherri tomatoes Figure 5.2) will most likely not benefit from such an analysis.

Task sequencing, in particular harvest sequencing, has been shown before to be practical for melon and apple harvesting [Mann et al., 2016; Zion et al., 2014; Edan et al., 2000; Edan et al., 1993; Edan et al., 1991]. The dynamic aspects described in Chapter 5 will be relevant to crops with fruit detected dynamically (e.g., melon locations can be tagged a priori and do not move but apples' locations will differ along harvest due to wind blowing, removal of other fruit). Since determining the approach direction can decrease cycle time, real-time algorithms for reachability should be developed in future research. They could be based on reachability cones (e.g., [Bloch, 2017]) which must be combined with real-time obstacle avoidance [Barth et al., 2016; Bac, 2015].

8.5 Application to other crops

Given a suitable crop and sensors the following steps should be advanced to benefit from the research developed in this thesis. The first step is to collect data with an autonomous system using the protocols developed in this thesis. The data collection should include all feasible conditions that might impact the visibility and detectability (e.g., angles, illumination conditions, crop growing directions, as detailed in Chapter 3.2). Once the data collection protocol is defined, and images are collected, a visibility and detectability analysis should be performed as outlined in Chapter 4 to generate guidelines to derive the best viewpoint/s. If the results are definitive and reproducible in different growing conditions then the guidelines should be defined statically, and the locations of sensors should be defined. If the detectability is dynamic, and changes as a function of environmental conditions, the dynamic sensing algorithm described in Chapter 5 should be adapted to the crop. The dynamic sensing algorithm should include a measure of information content, and actions to be made to increase it as a result of a previous sensing if needed. Finally, when the sensing strategy is defined, the number of cases in which occlusions cause loss of fruit in fruit approach should be identified using the methodology described in Chapter 7. If the number is high, the approach direction strategy should be evaluated, and humans in the loop should be considered along the protocols defined in this thesis. The overall optimization procedure should be measured in two main ways – harvesting cycle time optimization and detectability increase.

> *If we knew what it was we were doing, it would not be called research, would it?*
>
> **— Albert Einstein**

Bibliography

9

Ali, Osman, Bart Verlinden, & Dirk van Oudheusden (2009). “Infield logistics planning for crop-harvesting operations”. *Engineering Optimization* 41.2, pp. 183–197 (cit. on pp. 14, 70).

Al-Allaf, Omaima N (2014). “Review of face detection systems based artificial neural networks algorithms”. *The International Journal of Multimedia & Its Applications* 6.1, p. 1 (cit. on p. 12).

De-An, Zhao, Lv Jidong, Ji Wei, Zhang Ying, & Chen Yu (2011). “Design and control of an apple harvesting robot”. *Biosystems Engineering* 110.2, pp. 112–122 (cit. on pp. 11, 25).

Arad, Boaz, Polina Kurtser, Barnea Ehud, Ben Harel, Yael Edan, & Ohad Ben-Shahar (2018). “Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper harvesting robots”. *Submitted* (cit. on p. 12).

Arroyo, Javier, Maria Guijarro, & Gonzalo Pajares (2016). “An instance-based learning approach for thresholding in crop images under different outdoor conditions”. *Computers and Electronics in Agriculture* 127, pp. 669–679 (cit. on p. 12).

Bac, Wouter (2015). “Improving obstacle awareness for robotic harvesting of sweet-pepper”. PhD thesis. Wageningen University (cit. on pp. 11, 13, 32, 86, 87, 108, 109).

Bac, Wouter, Jochen Hemming, & Eldert van Henten (2014a). “Stem localization of sweet-pepper plants using the support wire as a visual cue”. *Computers and Electronics in Agriculture* 105, pp. 111–120 (cit. on pp. 87, 104, 108).

Bac, Wouter, Jochen Hemming, Barth van Tuijl, Ruud Barth, Ehud Wais, & Eldert van Henten (2017). “Performance evaluation of a harvesting robot for sweet pepper”. *Journal of Field Robotics* 34.6, pp. 1123–1139 (cit. on pp. 1, 3, 11, 13, 14, 87, 108).

Bac, Wouter, Tim Roorda, Roi Reshef, Sigal Berman, Jochen Hemming, & Eldert van Henten (2016). “Analysis of a motion planning problem for sweet-pepper harvesting in a dense obstacle environment”. *Biosystems Engineering* 146, pp. 85–97 (cit. on pp. 32, 87, 89).

Bac, Wouter, Eldert van Henten, Jochen Hemming, & Yael Edan (2014b). “Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead”. *Journal of Field Robotics* 31.6, pp. 888–911 (cit. on pp. 1, 3, 11–14, 24–27, 32, 51, 52, 60, 69, 70, 86, 87, 92, 104).

Bajcsy, Ruzena (1988). “Active perception”. *Proceedings of the IEEE* 76.8, pp. 966–1005 (cit. on pp. 8, 13, 52).

Bargoti, Suchet & James P Underwood (2017). “Image segmentation for fruit detection and yield estimation in apple orchards”. *Journal of Field Robotics* 34.6, pp. 1039–1060 (cit. on p. 12).

Barth, Ruud, Jochen Hemming, & Eldert van Henten (2016). “Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation”. *Biosystems Engineering* 146, pp. 71–84 (cit. on pp. 4, 8, 11–13, 26, 70, 76, 87, 90, 109).

Bechar, Avital & Yael Edan (2003). “Human-robot collaboration for improved target recognition of agricultural robots”. *Industrial Robot: An International Journal* 30.5, pp. 432–436 (cit. on pp. 87, 88).

Berenstein, Ron & Yael Edan (2012). “Human-Robot Cooperative Precision Spraying: Collaboration Levels and Optimization Function”. *IFAC Proceedings Volumes* 45.22, pp. 799–804 (cit. on pp. 87, 88).

Berenstein, Ron, Ohad Ben Shahar, Amir Shapiro, & Yael Edan (2010). “Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer”. *Intelligent Service Robotics* 3.4, pp. 233–243 (cit. on p. 66).

Blackmore, Simon (2016). “Towards robotic agriculture”. *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*. Vol. 9866. International Society for Optics and Photonics, p. 986603 (cit. on pp. 1, 7).

Blackmore, Simon, Henrik Have, & Spyros Fountas (2002). “A specification of behavioural requirements for an autonomous tractor”. *Automation Technology for Off-road Equipment, edited by Zhang, Q. ASAE Publication*, pp. 33–42 (cit. on pp. 27, 51).

Blanco, Jose-Luis, Francisco-Angel Moreno, & Javier Gonzalez (2009). “A collection of outdoor robotic datasets with centimeter-accuracy ground truth”. *Autonomous Robots* 27.4, p. 327 (cit. on p. 25).

Bloch, Victor (2017). “Agricultural Environment Characterization for Design of Optimal Robotic Armsr”. PhD thesis. Technion – Israel Institute of Technology (cit. on pp. 13, 87, 109).

Bloch, Victor, Avital Bechar, & Amir Degani (2017). “Development of an environment characterization methodology for optimal design of an agricultural robot”. *Industrial Robot: An International Journal* 44.1, pp. 94–103 (cit. on p. 11).

Bochtis, DD, CG Sørensen, & SG Vougioukas (2010). “Path planning for in-field navigation-aiding of service units”. *Computers and Electronics in Agriculture* 74.1, pp. 80–90 (cit. on p. 70).

Bochtis, DD & Claus G Sørensen (2009). “The vehicle routing problem in field logistics part I”. *Biosystems Engineering* 104.4, pp. 447–457 (cit. on pp. 14, 70).

– (2010). “The vehicle routing problem in field logistics: Part II”. *Biosystems Engineering* 105.2, pp. 180–188 (cit. on pp. 14, 70).

Bochtis, Dionysis, Hans Werner Griepentrog, Stavros Vougioukas, Patrizia Busato, Remigio Berruto, & K Zhou (2015). “Route planning for orchard operations”. *Computers and electronics in agriculture* 113, pp. 51–60 (cit. on p. 14).

Bontsema, Jan, Jochen Hemming, Erik Pekkeriet, Wouter Saeys, Yael Edan, Amir Shapiro, Marko Hočevar, Roberto Oberti, Manuel Armada, Heinz Ulbrich, et al. (2015). “CROPS: Clever robots for crops”. *Engineering and Technology Reference* 1.1 (cit. on pp. 12, 26, 49, 70, 86, 104).

Bulanon, D M, T F Burks, & V Alchanatis (2009). “Fruit visibility analysis for robotic citrus harvesting”. *Transactions of the ASABE* 52.1, pp. 277–283 (cit. on pp. 3, 7, 12, 13, 15, 24, 27, 33, 51, 52, 60, 109).

Chang, Wen-Chung (2007). “Precise positioning of binocular eye-to-hand robotic manipulators”. *Journal of Intelligent and Robotic Systems* 49.3, pp. 219–236 (cit. on pp. 4, 87).

Comba, L, P Gay, P Piccarolo, & D Ricauda Aimonino (2010). “Robotics and Automation for Crop Management : Trends and Perspective”. *International Conference Ragusa SHWA2010*, pp. 471–478 (cit. on p. 86).

Daly, Christopher, Mark P Widrlechner, Michael D Halbleib, Joseph I Smith, & Wayne P Gibson (2012). “Development of a new USDA plant hardiness zone map for the United States”. *Journal of Applied Meteorology and Climatology* 51.2, pp. 242–264 (cit. on p. 30).

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei (2009). “Imagenet: A large-scale hierarchical image database”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Ieee, pp. 248–255 (cit. on pp. 12, 25).

Dollar, Piotr, Christian Wojek, Bernt Schiele, & Pietro Perona (2012). “Pedestrian detection: An evaluation of the state of the art”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.4, pp. 743–761 (cit. on pp. 13, 25).

Duckett, Tom, Simon Pearson, Simon Blackmore, & Bruce Grieve (2018). “Agricultural Robotics: The Future of Robotic Agriculture”. *arXiv preprint arXiv:1806.06762* (cit. on p. 1).

Edan, Yael, B A Engel, & G E Miles (1993). “Intelligent control system simulation of an agricultural robot”. *Journal of Intelligent and Robotic Systems* 8.2, pp. 267–284 (cit. on pp. 11, 86, 109).
Edan, Yael, Tamar Flash, Uri M Peiper, Itzhak Shmulevich, & Yoav Sarig (1991). “Near-minimum-time task planning for fruit-picking robots”. *IEEE Transactions on Robotics and Automation* 7.1, pp. 48–56 (cit. on pp. 11, 15, 69, 70, 86, 109).
Edan, Yael & Gaines E Miles (1994). “Systems engineering of agricultural robot design”. *IEEE transactions on systems, man, and cybernetics* 24.8, pp. 1259–1265 (cit. on pp. 1, 4, 7, 11, 15).
Edan, Yael, Dima Rogozin, Tamar Flash, & Gaines E Miles (2000). “Robotic melon harvesting”. *IEEE Transactions on Robotics and Automation* 16.6, pp. 831–835 (cit. on pp. 4, 52, 109).
Edwards, Gareth TC, Jørgen Hinge, Nick Skou-Nielsen, Andrés Villa-Henriksen, Claus Aage Grøn Sørensen, & Ole Green (2017). “Route planning evaluation of a prototype optimised in-field route planner for neutral material flow agricultural operations”. *Biosystems Engineering* 153, pp. 149–157 (cit. on p. 14).
Eizicovits, Danny & Sigal Berman (2014). “Efficient sensory-grounded grasp pose quality mapping for gripper design and online grasp planning”. *Robotics and Autonomous Systems* 62.8, pp. 1208–1219 (cit. on pp. 11, 87).
Eizicovits, Danny, Bart van Tuijl, Sigal Berman, & Yael Edan (2016). “Integration of perception capabilities in gripper design using graspability maps”. *Biosystems Engineering* 146, pp. 98–113 (cit. on pp. 11, 27, 38, 87).
Elkoby, Zohar (2016). “Analysis of Human-Robot Harvesting Operations in Sweet Pepper Greenhouses”. MSc. Thesis. Beer Sheva, Israel: Ben-Gurion University of the Negev (cit. on pp. 3, 13, 53, 59, 63).
Elkoby, Zohar, van Bert Ooster, & Yael Edan (2014). “Simulation Analysis of Sweet Pepper Harvesting Operations”. *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World*, pp. 441–448 (cit. on pp. 3, 13, 53, 59, 63, 78).
Feng, Qingchun, Wei Zou, Pengfei Fan, Chunfeng Zhang, & Xiu Wang (2018). “Design and test of robotic harvesting system for cherry tomato”. *International Journal of Agricultural and Biological Engineering* 11.1, pp. 96–100 (cit. on p. 11).
Fleishman, Shachar, Daniel Cohen-Or, & Dani Lischinski (2000). “Automatic camera placement for image-based modeling”. *Computer Graphics Forum*. Vol. 19. 2. Wiley Online Library, pp. 101–110 (cit. on pp. 24, 27).
Fleiss, Joseph L & Jacob Cohen (1973). “The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability”. *Educational and Psychological Measurement* 33.3, pp. 613–619 (cit. on p. 96).
Foix Salmerón, Sergi, Guillem Alenyá Ribas, & Carme Torras (2011). “Towards plant monitoring through next best view”. *Artificial Intelligence Research and Development: Proceedings of the 14th International Conference of the Catalan Association for Artificial Intelligence*. IOS Press, pp. 101–109 (cit. on pp. 24, 27).
Foix, Sergi, Guillem Alenya, & Carme Torras (2015). “3D Sensor planning framework for leaf probing”. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, pp. 6501–6506 (cit. on pp. 27, 53).
Font, D, T Pallejà, M Tresanchez, M Teixidó, D Martinez, J Moreno, & J Palacín (2014). “Counting red grapes in vineyards by detecting specular spherical reflection peaks in RGB images obtained at night with artificial illumination”. *Computers and Electronics in Agriculture* 108, pp. 105–111 (cit. on p. 86).

Geiger, Andreas, Philip Lenz, & Raquel Urtasun (2012). “Are we ready for autonomous driving? the kitti vision benchmark suite”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 3354–3361 (cit. on pp. 25, 28).
Gendreau, Michel, Gilbert Laporte, & René Séguin (1996). “Stochastic vehicle routing”. *European Journal of Operational Research* 88.1, pp. 3–12 (cit. on p. 15).
Golden, Bruce L, Subramanian Raghavan, & Edward A Wasil (2008). The vehicle routing problem: latest advances and new challenges. Vol. 43. Springer Science and Business Media (cit. on p. 15).
Gongal, A, S Amatya, M Karkee, Q Zhang, & K Lewis (2015). “Sensors and systems for fruit detection and localization: A review”. *Computers and Electronics in Agriculture* 116, pp. 8–19 (cit. on pp. 12, 24, 86).
Guyon, Isabelle (1997). “A scaling law for the validation-set training-set size ratio”. *AT&T Bell Laboratories*, pp. 1–11 (cit. on p. 34).
Han, Kil-Su, Si-Chan Kim, Young Bum Lee, Sang Chul Kim, Dong Hyuk Im, Hong Ki Choi, & Heon Hwang (2012). “Strawberry harvesting robot for bench-type cultivation”. *Biosystems Engineering* 37.1, pp. 65–74 (cit. on pp. 1, 87).
Hansen, Eric A (1994). “Cost-Effective Sensing during Plan Execution.” *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, pp. 1029–1035 (cit. on p. 70).
Harel, Ben, Polina Kurtser, Liesbet van Herck, Yisrael Parmet, & Yael Edan (2016). “Sweet pepper maturity evaluation via multiple viewpoints color analyses”. *Proceedings, Conference on Agricultural Engineering (AgEng)* (cit. on p. 87).
Hayashi, Shigehiko, Katsunobu Ganno, Yukitsugu Ishii, & Itsuo Tanaka (2002). “Robotic harvesting system for eggplants”. *Japan Agricultural Research Quarterly: JARQ* 36.3, pp. 163–168 (cit. on pp. 11, 26).
Hayashi, Shigehiko, Kenta Shigematsu, Satoshi Yamamoto, Ken Kobayashi, Yasushi Kohno, Junzo Kamata, & Mitsutaka Kurita (2010). “Evaluation of a strawberry-harvesting robot in a field test”. *Biosystems Engineering* 105.2, pp. 160–171 (cit. on p. 11).
Hellström, Thomas & Ola Ringdahl (2013). “A software framework for agricultural and forestry robots”. en. *Industrial Robot: An International Journal* 40.1, pp. 20–26 (cit. on p. 91).
Hemming, Jochen, Wouter Bac, Barth van Tuijl, Ruud Barth, Jan Bontsema, Erik Pekkeriet, & Eldert van Henten (2014a). “A robot for harvesting sweet-pepper in greenhouses”. *International Conference of Agricultural Engineering* (cit. on pp. 1, 12, 14, 49, 60, 69).
Hemming, Jochen, Jos Ruizendaal, Jan Willem Hofstee, & Eldert van Henten (2014b). “Fruit detectability analysis for different camera positions in sweet-pepper”. *Sensors* 14.4, pp. 6032–6044 (cit. on pp. 3, 7, 12, 13, 15, 24, 25, 27, 33, 41, 51, 52, 55, 60, 69, 77, 86, 87).
Horizon 2020 SWEEPER (2015-2018). `http://www.sweeper-robot.eu` (cit. on pp. 9, 12).
Hosmer Jr, David W, Stanley Lemeshow, & Rodney X Sturdivant (2013). Applied Logistic Regression. Vol. 398. John Wiley & Sons (cit. on p. 93).
Jensen, Martin F, Dionysis Bochtis, & Claus G Sørensen (2015). “Coverage planning for capacitated field operations, part II: Optimisation”. *Biosystems Engineering* 139, pp. 149–164 (cit. on p. 14).
Jimenez, AR, R Ceres, & JL Pons (2000). “A survey of computer vision methods for locating fruit on trees”. *Transactions of the ASAE* 43.6, p. 1911 (cit. on p. 29).
Kamilaris, Andreas & Francesc X Prenafeta-Boldú (2018). “Deep learning in agriculture: A survey”. *Computers and Electronics in Agriculture* 147, pp. 70–90 (cit. on pp. 1, 3, 11–13, 27).
Kapach, Keren, Ehud Barnea, Rotem Mairon, Yael Edan, & Ohad Ben-Shahar (2012). “Computer vision for fruit harvesting robots–state of the art and challenges ahead”. *International*

Journal of Computational Vision and Robotics 3.1-2, pp. 4–34 (cit. on pp. 7, 11–13, 24, 27, 69, 86, 87, 104).

Kim, Hyun, Young-Jo Cho, & Sang-Rok Oh (2005). "CAMUS: A middleware supporting context-aware services for network-based robots". *Advanced Robotics and its Social Impacts, 2005. IEEE Workshop on*. IEEE, pp. 237–242 (cit. on pp. 1, 2).

Kitamura, S & K Oka (2005). "Recognition and cutting system of sweet pepper for picking robot in greenhouse horticulture". *IEEE International Conference Mechatronics and Automation, 2005*. Vol. 4, 1807–1812 Vol. 4 (cit. on p. 87).

Krainin, Michael, Peter Henry, Xiaofeng Ren, & Dieter Fox (2010). "Manipulator and object tracking for in hand model acquisition". *Proceedings, IEEE International Conference on Robots and Automation (ICRA)*. Anchorage, Alaska. (cit. on pp. 13, 26, 52).

Kuffner, James J & Steven M LaValle (2000). "RRT-connect: An efficient approach to single-query path planning". *Proceedings, IEEE International Conference on Robots and Automation (ICRA)*. Vol. 2. IEEE, pp. 995–1001 (cit. on p. 75).

Kurtser, Polina, Boaz Arad, Ohad Ben Shahar, Milan van Bree, Joep Moonen, Bart van Tujil, & Yael Edan (2016). "Robotic data acquisition of sweet pepper images for research and development". *The 5th Israeli Conference on Robotics, 13-14 April 2016. Hertzilya, Israel* (cit. on pp. 38, 95).

Kurtser, Polina & Yael Edan (2018a). "Statistical models for fruit detectability: spatial and temporal analyses of sweet peppers". *Biosystems Engineering* 171, pp. 272–289 (cit. on pp. 8, 51–53, 55, 57, 60, 72).

– (2018b). "The use of dynamic sensing strategies to improve detection for a pepper harvesting robot." *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (cit. on p. 8).

Landis, J Richard & Gary G Koch (1977). "The measurement of observer agreement for categorical data". *Biometrics*, pp. 159–174 (cit. on p. 96).

Laporte, Gilbert (1992). "The traveling salesman problem: An overview of exact and approximate algorithms". *European Journal of Operational Research* 59.2, pp. 231–247 (cit. on pp. 14, 70, 72).

Lass, Lawrence W & Timothy S Prather (2004). "Detecting the locations of Brazilian pepper trees in the Everglades with a hyperspectral sensor". *Weed Technology* 18.2, pp. 437–442 (cit. on pp. 52, 86).

Lawler, Eugene L (1985). "The traveling salesman problem: a guided tour of combinatorial optimization". *Wiley-Interscience Series in Discrete Mathematics* (cit. on pp. 14, 70, 72).

Lee, BS & UA Rosa (2006). "Development of a canopy volume reduction technique for easy assessment and harvesting of Valencia citrus fruits". *Transactions of the ASABE* 49.6, pp. 1695–1703 (cit. on p. 25).

Lehnert, Christopher, Andrew English, Christopher McCool, Adam W Tow, & Tristan Perez (2017). "Autonomous sweet pepper harvesting for protected cropping systems". *IEEE Robotics and Automation Letters* 2.2, pp. 872–879 (cit. on p. 11).

Lehnert, Christopher, Inkyu Sa, Christopher McCool, Ben Upcroft, & Tristan Perez (2016). "Sweet pepper pose detection and grasping for automated crop harvesting". *Proceedings, IEEE International Conference on Robots and Automation (ICRA)*. IEEE, pp. 2428–2434 (cit. on p. 87).

Lenstra, Jan Karel & AHG Rinnooy Kan (1981). "Complexity of vehicle routing and scheduling problems". *Networks* 11.2, pp. 221–227 (cit. on p. 70).

Lloyd, Stuart (1982). "Least squares quantization in PCM". *IEEE transactions on Information Theory* 28.2, pp. 129–137 (cit. on p. 53).

Luo, Lufeng, Hanjin Wen, Qinghua Lu, Haojie Huang, Weilin Chen, Xiangjun Zou, & Chenglin Wang (2018). “Collision-Free Path-Planning for Six-DOF Serial Harvesting Robot Based on Energy Optimal and Artificial Potential Field”. *Complexity* 2018 (cit. on p. 11).

MacKay, David J. C. (1992). “Information-Based Objective Functions for Active Data Selection”. *Neural Computation* 4.4, pp. 590–604 (cit. on pp. 8, 13, 52).

Mann, Moshe P, Boaz Zion, Dror Rubinstein, Raphael Linker, & Itzhak Shmulevich (2014). “Minimum time kinematic motions of a cartesian mobile manipulator for a fruit harvesting robot”. *Journal of Dynamic Systems, Measurement, and Control* 136.5, p. 051009 (cit. on p. 70).

Mann, Moshe P, Boaz Zion, Itzhak Shmulevich, Dror Rubinstein, & Raphael Linker (2016). “Combinatorial Optimization and Performance Analysis of a Multi-arm Cartesian Robotic Fruit Harvester—Extensions of Graph Coloring”. *Journal of Intelligent and Robotic Systems* 82.3-4, pp. 399–411 (cit. on pp. 14, 15, 69, 70, 86, 109).

Maver, Jasna & Ruzena Bajcsy (1993). “Occlusions as a Guide for Planning the Next View”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.5, pp. 417–433 (cit. on pp. 13, 24, 26).

McCool, Christopher, Inkyu Sa, Feras Dayoub, Christopher Lehnert, Tristan Perez, & Ben Upcroft (2016). “Visual detection of occluded crop: For automated harvesting”. *Proceedings, IEEE International Conference on Robots and Automation (ICRA)*, pp. 2506–2512 (cit. on pp. 1, 14, 24, 48, 69, 70, 87).

Mühler, Konrad, Mathias Neugebauer, Christian Tietjen, & Bernhard Preim (2007). “Viewpoint selection for intervention planning.” *EUROVIS07 Proceedings of the 9th Joint Eurographics / IEEE VGTC conference on Visualization*, pp. 267–274 (cit. on pp. 13, 26, 52).

Murphy, Kevin, Brett Myors, & Allen Wolach (2014). Statistical power analysis: A simple and general model for traditional and modern hypothesis tests. Routledge (cit. on p. 34).

Murphy, Robin (2000). Introduction to AI robotics. MIT press (cit. on pp. 1, 2).

Olawsky, Duane, Kurt Krebsbach, & Maria Gini (1993). *An analysis of sensor-based task planning*. Tech. rep. Technical Report 93-94, Computer Science Department, University of Minnesota, Minneapolis (Minnesota) (cit. on p. 70).

Orfanou, A, P Busato, DD Bochtis, G Edwards, D Pavlou, Claus Grøn Sørensen, & R Berruto (2013). “Scheduling for machinery fleets in biomass multiple-field operations”. *Computers and Electronics in Agriculture* 94, pp. 12–19 (cit. on p. 70).

Ostovar, Ahmad, Ola Ringdahl, & Thomas Hellström (2018). “Adaptive Image Thresholding of Yellow Peppers for a Harvesting Robot”. *Robotics* 7.1, p. 11 (cit. on p. 12).

Pekkeriet, EJ (2011). “CROPS project deliverable 12.1: Economic viability for each application”. *Wageningen, The Netherlands: Wageningen UR Greenhouse Horticulture* (cit. on pp. 13, 60).

Philbin, James, Ondrej Chum, Michael Isard, Josef Sivic, & Andrew Zisserman (2007). “Object retrieval with large vocabularies and fast spatial matching”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1–8 (cit. on p. 15).

Pillac, Victor, Michel Gendreau, Christelle Guéret, & Andrés L Medaglia (2013). “A review of dynamic vehicle routing problems”. *European Journal of Operational Research* 225.1, pp. 1–11 (cit. on p. 15).

Pitla, Santosh K (2018). “Agricultural Robotics”. *Advances in Agricultural Machinery and Technologies*. CRC Press, pp. 157–177 (cit. on p. 1).

Plebe, Alessio & Giorgio Grasso (2001). “Localization of spherical fruits for robotic harvesting”. *Machine Vision and Applications* 13.2, pp. 70–79 (cit. on p. 25).

Privault, Nicolas (2013). “Spatial Poisson Processes”. *Understanding Markov Chains*. Springer, pp. 225–239 (cit. on p. 58).

Reed, Michael K & Peter K Allen (2000). “Constraint-based sensor planning for scene modeling”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.12, pp. 1460–1467 (cit. on pp. 24, 27, 52).

Ringdahl, Ola, Polina Kurtser, Ruud Barth, & Yael Edan (2016). “Operational flow of an autonomous sweetpepper harvesting robot”. *The 5th Israeli Conference on Robotics, 13-14 April 2016. Hertzilya, Israel* (cit. on p. 91).

Ringdahl, Ola, Polina Kurtser, & Yael Edan (2017). “Strategies for selecting best approach direction for a sweet-pepper harvesting robot”. *Towards Autonomous Robotic Systems: 18th Annual Conference*. Guildford, UK: Springer, pp. 516–525 (cit. on pp. 9, 49, 57, 88, 91–93, 95, 102).

– (2018). “Evaluation of approach strategies for harvesting robots: Case study of sweet pepper harvesting”. *Journal of Intelligent and Robotic Systems*, pp. 1–16 (cit. on pp. 9, 52, 57).

Rong, Dian, Yibin Ying, & Xiuqin Rao (2017). “Embedded vision detection of defective orange by fast adaptive lightness correction algorithm”. *Computers and Electronics in Agriculture* 138, pp. 48–59 (cit. on p. 12).

Rosenbaum, David A., Rajal G. Cohen, Ruud G. J. Meulenbroek, & Jonathan Vaughan (2006). “Plans for Grasping Objects”. *Motor Control and Learning*. Boston: Kluwer Academic Publishers, pp. 9–25 (cit. on p. 87).

Rosenkrantz, Daniel J, Richard E Stearns, & Philip M Lewis II (1977). “An analysis of several heuristics for the traveling salesman problem”. *SIAM Journal on Computing* 6.3, pp. 563–581 (cit. on p. 14).

Russell, Bryan C, Antonio Torralba, Kevin P Murphy, & William T Freeman (2008). “LabelMe: a database and web-based tool for image annotation”. *International Journal of Computer Vision* 77.1, pp. 157–173 (cit. on pp. 7, 12, 13, 15, 25, 51).

Rylski, I & M Spigelman (1986). “Effect of shading on plant development, yield and fruit quality of sweet pepper grown under conditions of high temperature and radlation”. *Scientia Horticulturae* 29.1-2, pp. 31–35 (cit. on p. 30).

Sa, Inkyu, Zongyuan Ge, Feras Dayoub, Ben Upcroft, Tristan Perez, & Chris McCool (2016). “Deepfruits: A fruit detection system using deep neural networks”. *Sensors* 16.8, p. 1222 (cit. on pp. 1, 7, 11, 12, 15, 24, 34, 48, 69, 72, 87).

Sakia, R M (1992). “The Box-Cox Transformation Technique: A Review”. *Journal of the Royal Statistical Society. Series D (The Statistician)* 41.2, pp. 169–178 (cit. on p. 93).

Sakthivel, K, R Nallusamy, & C Kavitha (2015). “Color Image Segmentation Using SVM Pixel Classification Image”. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering* 8.10, pp. 1919–1925 (cit. on p. 12).

Sarig, Y (1993). “Robotics of fruit harvesting: A state-of-the-art review”. *Journal of Agricultural Engineering Research* 54.4, pp. 265–280 (cit. on pp. 1, 11).

Shamshiri, Redmond Ramin, Cornelia Weltzien, Ibrahim A Hameed, Ian J Yule, Tony E Grift, Siva K Balasundram, Lenka Pitonakova, Desa Ahmad, & Girish Chowdhary (2018). “Research and development in agricultural robotics: A perspective of digital farming”. *International Journal of Agricultural and Biological Engineering* 11.4, pp. 1–14 (cit. on p. 1).

Shmmala, Faten Abu & Wesam Ashour (2013). “Color based image segmentation using different versions of k-means in two spaces”. *Global Advanced Research Journal of Engineering, Technology and Innovation* 1.9, pp. 30–41 (cit. on p. 12).

Sites, Peter W & Michael J Delwiche (1988). “Computer vision to locate fruit on a tree”. *Transactions of the ASAE* 31.1, pp. 257–0265 (cit. on p. 25).

Suh, Hyun K, Jan Willem Hofstee, & Eldert van Henten (2018). “Improved vegetation segmentation with ground shadow removal using an HDR camera”. *Precision Agriculture* 19.2, pp. 218–237 (cit. on p. 86).

Sun, Jian, Sing Bing Kang, Zong-Ben Xu, Xiaoou Tang, & Heung-Yeung Shum (2007). "Flash cut: Foreground extraction with flash and no-flash image pairs". *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1–8 (cit. on pp. 29, 49).
Szeliski, Richard (2010). Computer vision: algorithms and applications. Springer Science & Business Media (cit. on p. 15).
Tadesse, Teshome, Errol W Hewett, Michael A Nichols, & Keith J Fisher (2002). "Changes in physicochemical attributes of sweet pepper cv. Domino during fruit growth and development". *Scientia Horticulturae* 93.2, pp. 91–103 (cit. on p. 92).
Tang, Jing-Lei, Xiao-Qian Chen, Rong-Hui Miao, & Dong Wang (2016). "Weed detection using image processing under different illumination for site-specific areas spraying". *Computers and Electronics in Agriculture* 122, pp. 103–111 (cit. on p. 86).
Tang, Xiuying, Tiezhong Zhang, Ling Liu, Dan Xiao, & Yizhe Chen (2009). "A new robot system for harvesting cucumber". *American Society of Agricultural and Biological Engineers Annual International Meeting Reno, Nevada*, p. 1 (cit. on p. 26).
Tanigaki, Kanae, Tateshi Fujiura, Akira Akase, & Junichi Imagawa (2008). "Cherry-harvesting robot". *Computers and Electronics in Agriculture* 63.1, pp. 65–72 (cit. on p. 11).
Tarabanis, Konstantinos A, Peter K Allen, & Roger Y Tsai (1995). "A survey of sensor planning in computer vision". *IEEE Transactions on Robotics and Automation* 11.1, pp. 86–104 (cit. on pp. 26, 27).
Tian, Lei F & David C Slaughter (1998). "Environmentally adaptive segmentation algorithm for outdoor image segmentation". *Computers and Electronics in Agriculture* 21.3, pp. 153–168 (cit. on p. 29).
Urrea, Claudio & José Muñoz (2015). "Path tracking of mobile robot in crops". *Journal of Intelligent and Robotic Systems* 80.2, pp. 193–205 (cit. on p. 86).
van Henten, EJ, DA van't Slot, CWJ Hol, & LG van Willigenburg (2009). "Optimal manipulator design for a cucumber harvesting robot". *Computers and electronics in agriculture* 65.2, pp. 247–257 (cit. on p. 11).
van Henten, Eldert, Jochen Hemming, Barth van Tuijl, JG Kornet, J Meuleman, Jan Bontsema, & EA van Os (2002). "An autonomous robot for harvesting cucumbers in greenhouses". *Autonomous Robots* 13.3, pp. 241–258 (cit. on pp. 1, 11, 26).
van Henten, Eldert, Barth van van Tuijl, Jochen Hemming, JG Kornet, Jan Bontsema, & EA van Os (2003). "Field test of an autonomous cucumber picking robot". *Biosystems Engineering* 86.3, pp. 305–313 (cit. on p. 11).
van Houwelingen, JC & S Le Cessie (1990). "Predictive value of statistical models". *Statistics in Medicine* 9.11, pp. 1303–1325 (cit. on p. 34).
Vázquez, Pere-Pau, Miquel Feixas, Mateu Sbert, & Wolfgang Heidrich (2001). "Viewpoint selection using viewpoint entropy." *VMV*. Vol. 1, pp. 273–280 (cit. on pp. 8, 13, 26, 52).
Vázquez, Pere-Pau, Miquel Feixas, Mateu Sbert, & Antoni Llobet (2002). "Viewpoint entropy: a new tool for obtaining good views of molecules". *ACM International Conference Proceeding Series*. Vol. 22, pp. 183–188 (cit. on pp. 8, 13, 52).
Vitzrabin, Efi & Yael Edan (2016a). "Adaptive thresholding with fusion using a RGBD sensor for red sweet-pepper detection". *Biosystems Engineering* 146, pp. 45–56 (cit. on pp. 3, 12, 27, 51, 65, 86, 87).
– (2016b). "Changing task objectives for improved sweet pepper detection for robotic harvesting". *IEEE Robotics and Automation Letters* 1.1, pp. 578–584 (cit. on pp. 12, 25).
Vougioukas, Stavros G (2012). "A distributed control framework for motion coordination of teams of autonomous agricultural vehicles". *Biosystems Engineering* 113.3, pp. 284–297 (cit. on p. 14).

Wang, Jianlun, Jianlei He, Yu Han, Changqi Ouyang, & Daoliang Li (2013). “An adaptive thresholding algorithm of field leaf image”. *Computers and Electronics in Agriculture* 96, pp. 23–39 (cit. on p. 12).
Willmott, Cort J (1981). “On the validation of models”. *Physical geography* 2.2, pp. 184–194 (cit. on pp. 34, 37).
Winkelmann, Rainer & Klaus F Zimmermann (1995). “Recent developments in count data modelling: theory and application”. *Journal of Economic Surveys* 9.1, pp. 1–24 (cit. on p. 36).
Wright, Sewall (1921). “Correlation and causation”. *Journal of Agricultural Research* 20.7, pp. 557–585 (cit. on p. 34).
Yoshida, Jituo, Shigeaki Okuyama, & Hiroshi Suzuki (1985). *Fruit harvesting apparatus with television camera and monitor*. US Patent 4,519,193 (cit. on pp. 52, 86).
Yuan, Su, Yang Lei, Song Xin, & Li Bing (2016). “Design and Experiment of Intelligent Mobile Apple Picking Robot”. *Journal of Agricultural Mechanization Research* 1, p. 035 (cit. on p. 11).
Zabulis, Xenophon & Kostas Daniilidis (2004). “Multi-Camera Reconstruction based on Surface Normal Estimation and Best Viewpoint Selection”. *Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium*. IEEE Computer Society, pp. 733–740 (cit. on pp. 13, 26).
Zemmour, Elie, Polina Kurtser, & Yael Edan (2017). “Dynamic thresholding algorithm for robotic apple detection”. *Autonomous Robot Systems and Competitions (ICARSC), 2017 IEEE International Conference on*. IEEE, pp. 240–246 (cit. on pp. 12, 65, 87).
Zhao, Yuanshen, Liang Gong, Yixiang Huang, & Chengliang Liu (2016). “A review of key techniques of vision-based control for harvesting robot”. *Computers and Electronics in Agriculture* 127, pp. 311–323 (cit. on pp. 11, 12).
Zheng, Liying, Jingtao Zhang, & Qianyu Wang (2009). “Mean-shift-based color segmentation of images containing green vegetation”. *Computers and Electronics in Agriculture* 65.1, pp. 93–98 (cit. on p. 12).
Zion, B, M Mann, D Levin, A Shilo, D Rubinstein, & I Shmulevich (2014). “Harvest-order planning for a multiarm robotic harvester”. *Computers and Electronics in Agriculture* 103, pp. 75–81 (cit. on pp. 14, 15, 69, 70, 109).

Appendixes

10

10.1 List of software and databases

The following software packages were developed in this thesis:

- Joint detectability labeling user interface (MATLAB) - Chapter 4
- Robotic viewpoint acquistion protocol software (ROS,C++)- Chapters 3.2,4
- Dynamic sensing algorithm implementation (MATLAB) - Chapter 5
- TSP simulation (MATLAB,V-Rep) - Chapter 6
- TSP laboratory experiment (Java) - Chapter 6
- Approach strategies implementation (ROS,C++) - Chapter 7
- Statistical modeling of viewpoints (SPSS,EXCEL) - Chapter 4
- Statistical modeling for prediction of target in the scene for dynamic viewpoint (SPSS,R,EXCEL) - Chapter 5
- Statistical analysis of questionnaires responses (SPSS,EXCEL)- Chapter 7

The following databases were acquired and are available for future research:

- Viewpoint databases (1-3), described and used in Chapter 4. DB#1 is available to the public.
- Viewpoint for dynamic sensing database, used in Chapter 5.
- Pepper locations database, used in Chapters 6 and 7.

10.2 SWEEPER public deliverable

SWEEPER

Sweet Pepper Harvesting Robot

GRANT AGREEMENT Number 644313

Deliverable number: 5.1
Deliverable title: Image database #1

WP number: 5
Lead beneficiary: BGU
Authors: Boaz Arad, Polina Kurtser, Yael Edan, Ohad Ben-Shahar

Contributors: Ruud Barth, Bart van Tuijl, Jochen Hemming, , Milan van Bree, Joep Moonen, Richard Overkamp

Nature: R
Dissemination level: PU
Delivery date: M7

Executive Summary

During two visit to the grower's greenhouse (July and Sep 2015), the first set of field data was collected and organized. While deliverable 5.1 is the dataset itself, this short report describes its design and content, and how to publically access it.

Contents

Executive Summary 2
1. Introduction 4
2. Dataset design and protocol 4
3. Dataset Content 6
4. Public web access and graphical user interface 7
Appendix A: 9

1. Introduction

Sweeper detection and localization algorithms are designed to be data-driven. To facilitate this approach, the Sweeper research plan includes no less than 4 data collection sessions to serve algorithm design for both the basic and advance system. In July and Sep 2015, teams from BGU, DLO, and Irmato met and collaborated in the Grower's greenhouse, to make initial sensor evaluation (in July) and then run the first systematic data collection (in Sep) that resulted in the first dataset of sweet pepper greenhouse scenes as eventually acquired by the Sweeper robot. According to the research plan, this dataset, as well as the forthcoming ones, should be made public both for the Sweeper community and the research community in general. The rest of this short document describes the content of this dataset and how to access it.

2. Dataset design and protocol

The Sweeper robot is likely to observe the sweet pepper plant from various angles and distances (as much as the space between aisles permits). In addition, illumination conditions may vary from direct sun light to complete darkness (during night time). To facilitate data collection under all these conditions we designed an acquisition protocol that utilizes the selected sensors, the selected sweeper manipulator, an available artificial illumination sources, and custom-made software, to collect data in the following way:

- Sensors were mounted on the tip of the manipulator (Fanuc LR mate 200iD, 900mm 7L version) that was programmed to move between 15 predefined configurations that cover 5 viewpoints at each of 3 distances from the plant. Since on the ground it was found that the furthest and highest viewpoint pushes the limits of the arm, that viewpoint was discarded, leaving 14 viewpoints for each scene.
- The manipulator itself was positioned on a lift that was manually moved along the aisle and lifted to the proper height to face scenes with sweet peppers.
- At each such view point, an image was taken from the RGB-D Fotonic sensor, and from the iDS camera. Furthermore, images from the latter were taken both under natural illumination, and under strobed artificial illumination.
- Upon completion of all viewpoints, the robot switched to a homing position, the platform was moved to a new place to face a new scene, and the entire sequence of operation restarted.

Figure 1 shows the sensory rig and the robot in the aisle.

Figure 1: The sensory rig on the top of the robotic manipulator are mounted on the platform in the greenhouse

Note that all RGB image from the iDS and Fotonics were taken under custom-made automatic exposure control. The automatic exposure mechanism was designed to adjust camera exposure interval in order to maximally match the resultant histogram to a desired canonical structure (that was measured from a large set of images that were judged “good” by a human observer and were not over or under saturated). While the present procedure only attempted to optimize histogram peak position, future version will try to optimize the entire histogram and consider regions of interest other than the center. The canonical histogram used is shown in Figure 2.

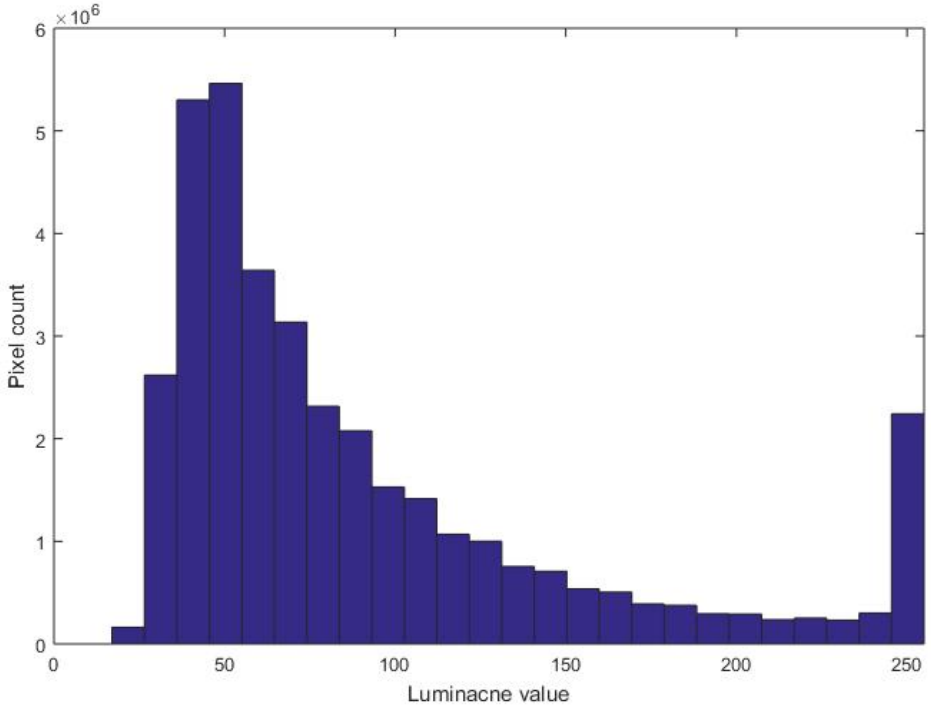


Figure 2: A canonical histogram that the auto exposure procedure was attempting to obtain by adjusting the exposure time of the camera. The present version of the auto exposure mechanisms attempted to adjust exposure time in order to obtain an image histogram with peak position that matches the canonical histogram. Future version will attempt to optimize additional features of the histogram.

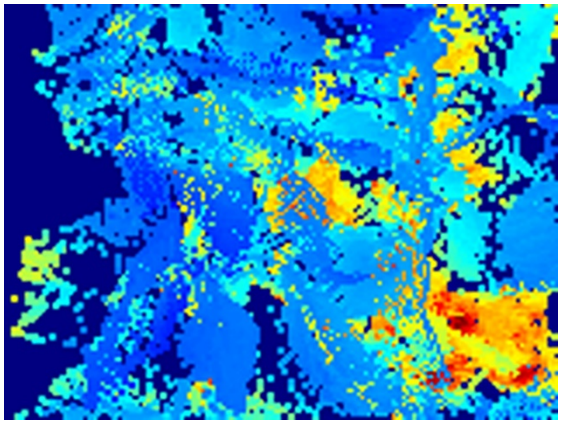
3. Dataset Content

Given the mechanisms and protocol as above, the data collection session included 1.5 days of collection with the robotic arm, including one night session. More specifically, the first Sweeper datasets includes

- A total of 43 scenes, each constituting 14 consistent viewpoints.
- At each viewpoint, 4 images are available (see Figure 3)
    - An RGB image from the Fotonic camera
    - A registered depth image from the Fotonic camera
    - An RGB image from the iDS camera (not registered with the Fotonics) under natural illumination
    - An RGB image from the iDS camera under artificial strobed light.
- Of the 43 scenes
    - 8 were taken on a cloudy day in the afternoon
    - 2 were taken during night time

- 14 were taken under direct sun light from behind the fruit (and thus directly into the lens)
- 19 were taken with the sun behind (and above) the sensors.

Fotonic RGB



Fotonic Depth

iDS RGB natural light

iDS RGB Strobed light

Figure 3: One sample of the 4 images taken from a single viewpoint.

4. Public web access and graphical user interface

All data of the first Sweeper DB are available publically through a web interface at the following URL:

http://www.cs.bgu.ac.il/~icvl/lab_projects/agrovision/DB/Sweeper01/

A snapshot of the main screen and the intuitive user interface is shown in Fig. 4. This web interface allows interactive browsing through the dataset, and it provides downloading features of a single image, a single image set (from a given viewpoint), or the entire dataset in one click of a button.

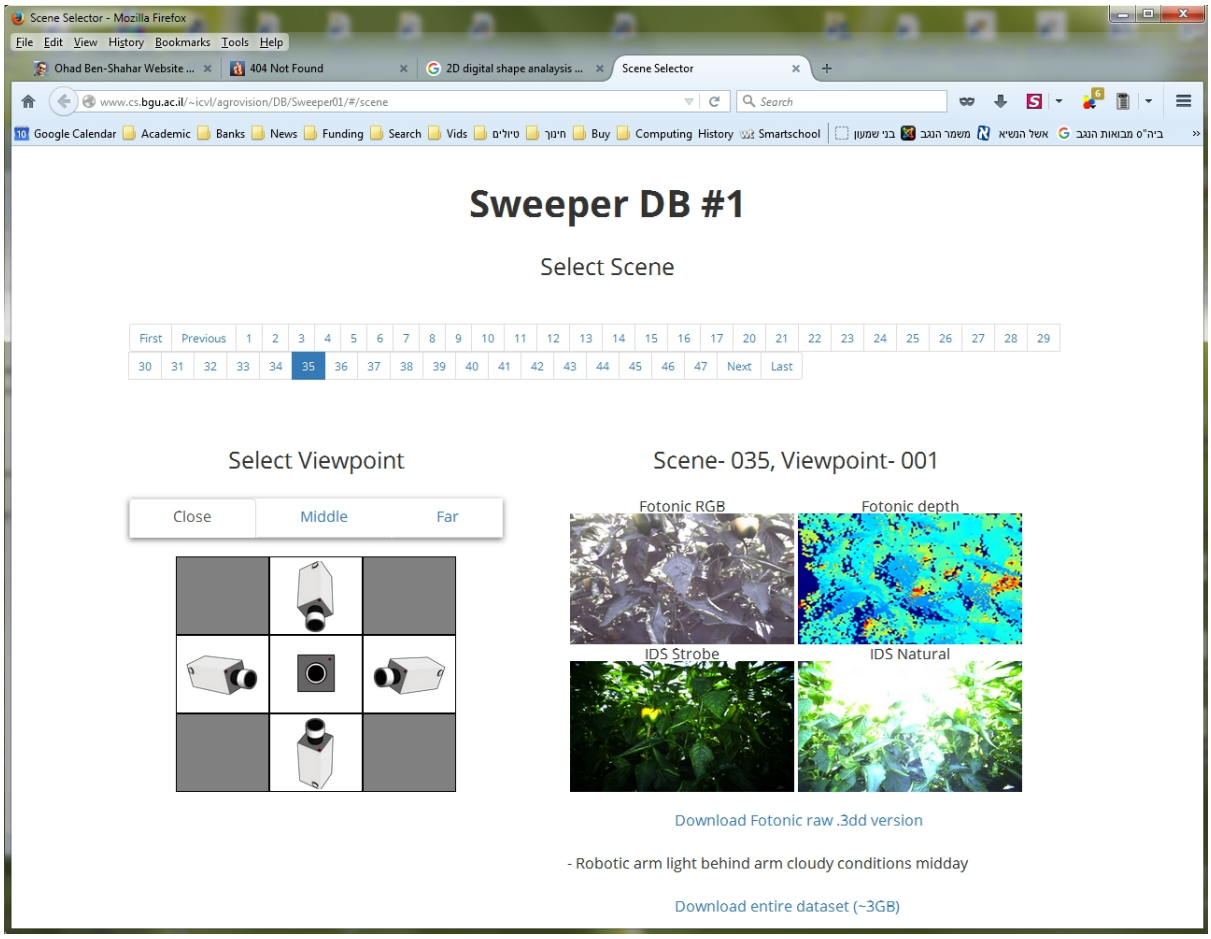


Figure 4: A snapshot of the interactive web interface to the first Sweeper DB.

Appendix A: Backup images

The robotic manipulator arrived at the greenhouse and due to some technical problems it was set up by Irmato for the experiment relatively late, leaving essentially one day for systematic data collection. In order to ensure that data is collected even if the manipulator does not work, an earlier effort of data collection employed a manual rig as shown in Figure 2. This rig (made by DLO's Bart van Tuijl) was placed on the platform, its viewpoint fixed, and images were acquired as the platform was moved along the aisle. Thus, unlike the data described above, each scene acquired included only a single set of images. Overall, 320 scenes were imaged this way (each constituting a set of the sort shown in Fig. 3) and this backup data is available from BGU upon request.

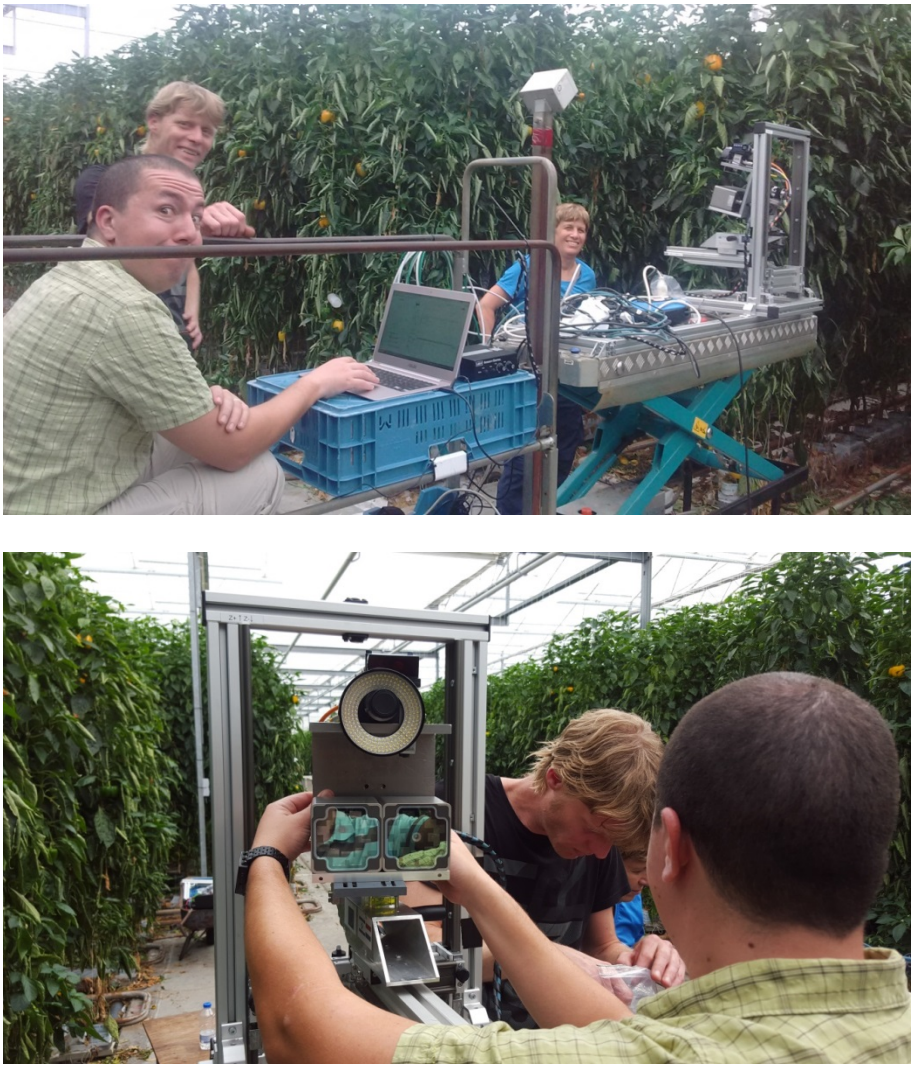


Figure 2: The sensory rig on the top of the manual rig – a contingency setup that was used before the robotic arm was utilized.

10.3 Approach direction - conference proceeding

Strategies for Selecting Best Approach Direction for a Sweet-Pepper Harvesting Robot

Ola Ringdahl[1], Polina Kurtser[2(✉)], and Yael Edan[2]

[1] Department of Computing Science,
Umeå University, 901 87 Umeå, Sweden
[2] Department of Industrial Engineering and Management,
Ben-Gurion University of the Negev, P.O.B. 653, 8410501 Beer Sheva, Israel
kurtser@post.bgu.ac.il

Abstract. An autonomous sweet pepper harvesting robot must perform several tasks to successfully harvest a fruit. Due to the highly unstructured environment in which the robot operates and the presence of occlusions, the current challenges are to improve the detection rate and lower the risk of losing sight of the fruit while approaching the fruit for harvest. Therefore, it is crucial to choose the best approach direction with least occlusion from obstacles.

The value of ideal information regarding the best approach direction was evaluated by comparing it to a method attempting several directions until successful harvesting is performed. A laboratory experiment was conducted on artificial sweet pepper plants using a system based on eye-in-hand configuration comprising a 6DOF robotic manipulator equipped with an RGB camera. The performance is evaluated in laboratorial conditions using both descriptive statistics of the average harvesting times and harvesting success as well as regression models. The results show roughly 40–45% increase in average harvest time when no a-priori information of the correct harvesting direction is available with a nearly linear increase in overall harvesting time for each failed harvesting attempt. The variability of the harvesting times grows with the number of approaches required, causing lower ability to predict them.

Tests show that occlusion of the front of the peppers significantly impacts the harvesting times. The major reason for this is the limited workspace of the robot often making the paths to positions to the side of the peppers significantly longer than to positions in front of the fruit which is more open.

1 Introduction

Due to the lack of skilled workforce and increasing labor costs, advanced automation is required for greenhouse production systems [1]. Despite intensive R&D on harvesting robots, there are no commercial harvesting robots for sweet peppers [2, 3]. Robotic harvesting of sweet peppers includes several tasks: detecting the fruit, approaching it, deciding whether the fruit is ripe, and finally detaching the fruit from the stem [4, 5]. The major limitation most commonly tackled today is the non-optimal detection rates; Bac et al. [3] reported state of the art being 85% in their 2014 review. Viewpoints analyses in harvesting robotics indicate that only 60% of the fruit can be detected from a single detection direction [6]. Therefore, current research focuses on detection

Y. Gao et al. (Eds.): TAROS 2017, LNAI 10454, pp. 516–525, 2017.
DOI: 10.1007/978-3-319-64107-2_41

algorithm development [3, 6–8]. Another challenge often described in the literature is the task of how to grasp a fruit, due to the limitations of available robotic grippers and the inherent difficulties of grasp planning [9, 10]. Eizicovits and Berman [10] developed geometry-based grasp quality measures based on 3D point cloud to determine the best grasping pose of different objects, including sweet peppers. This kind of solution depends on detailed 3D sensor information of the object [11] which is very difficult to achieve in dense greenhouse environments. These environments have an unstructured and dynamic nature [12]: fruits have a high inherent variability in size, shape, texture, and location; in addition, occlusion and variable illumination conditions significantly influence the detection performance. Given the complexity of both detection and grasp planning tasks, approaching the correct fruit pose must be done dynamically, taking into account obstacles such as stems and leaves. The most common way to do this is visual servoing, i.e. using eye-in-hand sensing to guide the robot towards the fruit by always keeping it in the center of the image [13]. When using this method, it is crucial to choose the best approach direction with least occlusion from leaves and other obstacles to maximize the chance for the visual servoing to reach the desired grasping pose. This research focuses on measuring the value of ideal information regarding the best approach direction for successful visual servoing, compared to a method using a search pattern to find the best direction.

2 Methods

A 6DOF robotic manipulator Fanuc LR Mate 200iD equipped with an eye-in hand iDS Ui-5250RE RGB camera and a Sick DT20HI displacement measurement laser sensor was placed in-front of an artificial plastic pepper crop with yellow plastic fruits and green leaves (Fig. 1). The workflow of the robot was implemented using a generic software framework for development of agricultural and forestry robots [14]. The framework is constructed with a hybrid robot architecture, using a state machine implementing a flowchart as described by Ringdahl et al. [15].

Fig. 1. The experimental setup consisted of a robotic harvester in front of an artificial crop.

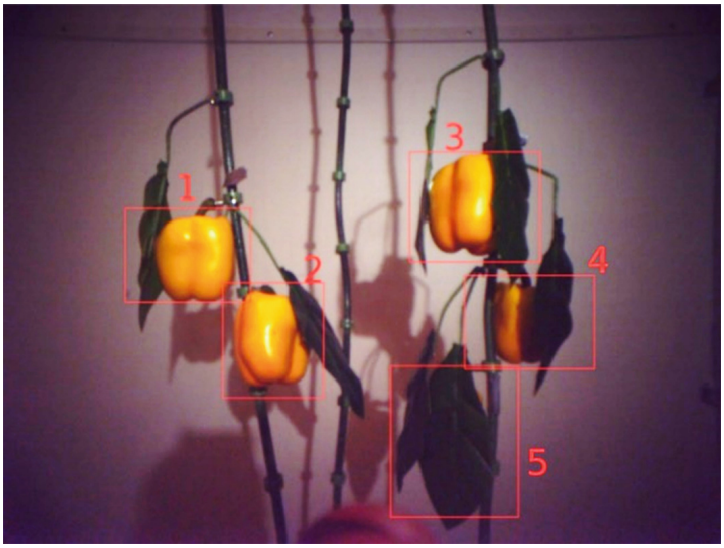


Fig. 2. An overview image taken from the robot's camera looking at a laboratorial scene with 5 peppers on two stems covered by leaves.

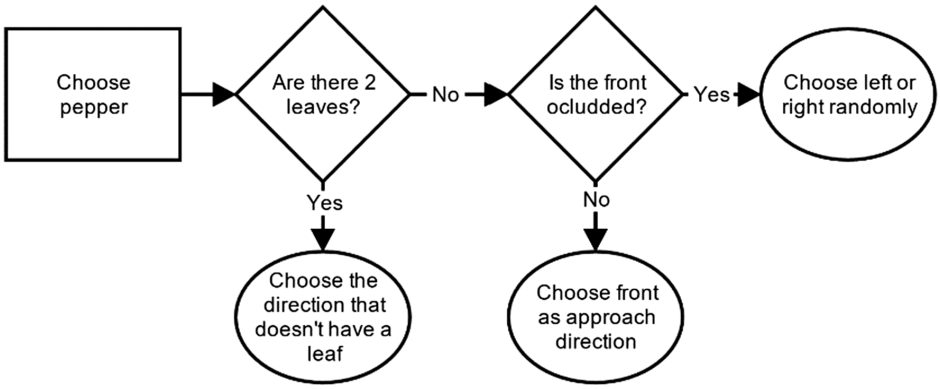


Fig. 3. Decision flowchart for manually selecting the optimal approach direction to a pepper.

A scene consisting of five plastic fruits placed at different locations on two artificial stems was setup before each experiment. The number of fruits were set to 5 to be similar to an actual sweet pepper plant, the right stem had three fruits, the left had two fruits. Each fruit had one or two leaves placed on different side (left/front/right) of it to create occlusion. An example of an overview image taken by the robot can be seen in Fig. 2. For each fruit the best fit harvesting approach, defined as the "optimal" harvesting approach direction, was set as the angle from either left (−45°), front (0°), or right (45°) where the target was least occluded, was noted manually. Figure 3 shows a flowchart describing the decision process for the manual selection.

2.1 Harvesting Scenarios

Two harvesting scenarios were tested. The first scenario, the *full a-priori knowledge scenario*, represents the ground-truth where both position $P_i(x_i, y_i, z_i)$ and approach

direction θ_i^* are known for each fruit i. The harvesting cycle consists of approaching a pre-defined overview waypoint $W_0(x, y, z)$, and then selecting each target fruit in order from the list of positions and optimal approach directions of all fruits. The control unit then calculates the path of the robotic manipulator to a waypoint $W_i(x, y, z)$, positioned at a defined distance from fruit i with respect to the optimal harvesting approach direction and position $\left(x_i, y_i, z_i, \theta_i^*\right)$. After reaching the waypoint, a visual servo procedure based on color blob detection and distance measurements received from the laser guides the manipulator towards the target until the end-effector touches the fruit. If the manipulator reaches the target fruit, the harvest of that fruit is marked as successful and the path to the next waypoint is then calculated. In case the fruit was not found or was lost from view while in visual servo, the harvest of the fruit is marked as failed and the path to the next waypoint is calculated. The cycle ends when all fruits have been attempted to be approached. The left part of Fig. 4 shows a flowchart of this harvesting scenario.

The second scenario, the *auto approach direction search scenario*, is a variation of the ground-truth scenario in which the optimal approach direction θ_i^* is unknown, and therefore must be searched from a list of predefined possible approach directions $\theta_1..\theta_k$. For each target fruit i and possible approach direction θ_j the control unit calculates the path of the robotic manipulator to a waypoint $W_{ij}(x, y, z)$ positioned at a defined distance from the target fruit with respect to θ_j until the harvest of the fruit is marked as successful or sight of the fruit is lost. If successful, the path to the waypoint W_{ij} for fruit $i+1$ and θ_1 is calculated. If the fruit was lost during visual servoing, the next approach direction θ_{j+1} is selected. In case all approach directions $\theta_1..\theta_k$ were attempted without being able to reach the fruit, the harvest of the target fruit is marked as failed and the path to the waypoint W_{ij} for fruit $i+1$ and θ_1 is calculated. The right part of Fig. 4 shows a flowchart of this harvesting scenario.

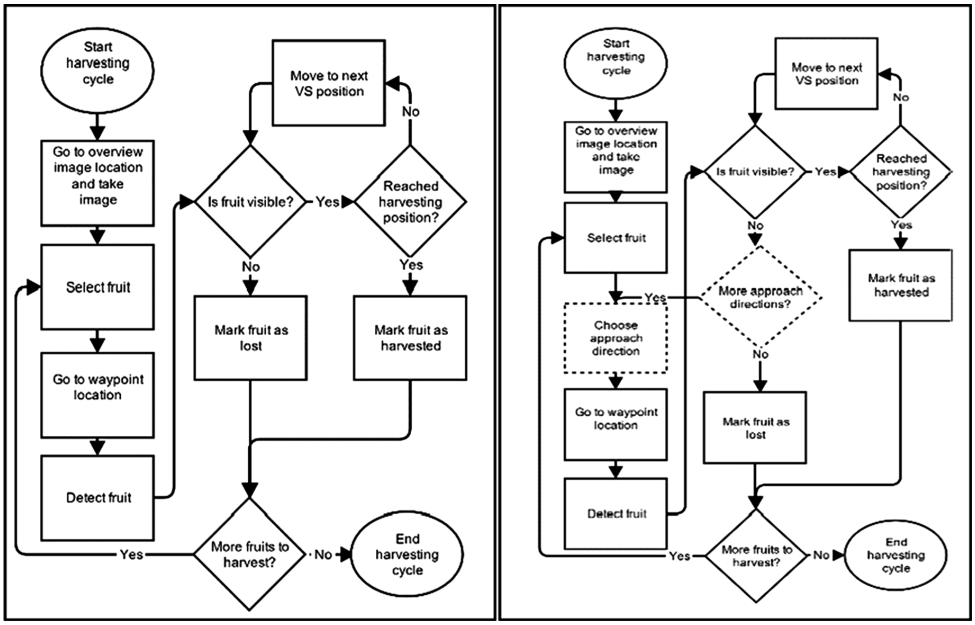


Fig. 4. Flowchart describing the two different harvesting scenarios. Left: auto approach direction search scenario. Right: full a-priori knowledge scenario (differences marked with dashed lines).

2.2 Experimental Protocol

Six laboratory scenes with different leaves and optimal approach directions were set up as defined in Table 1. The pose of each pepper was measured by manually moving the robotic arm in the desired approach direction into the position where the gripper touched the fruit, as seen in Fig. 5.

Table 1. Six scenes with different configurations for leaf (L = left, F = front, R = right) and approach direction (−45°, 0°, 45°).

Scene	Pepper 1				Pepper 2				Pepper 3				Pepper 4				Pepper 5			
	L	F	R		L	F	R		L	F	R		L	F	R		L	F	R	
1	×	×		45	×		×	0	×			0		×		−45			×	0
2	×		×	0		×	×	−45		×		−45	×			0			×	0
3	×		×	0		×	×	−45	×			0			×	0		×		−45
4	×			0			×	0	×		×	0		×		−45	×	×		45
5			×	0		×		45	×		×	0	×			0	×	×		45
6	×			0			×	0	×		×	0		×		−45	×	×		45

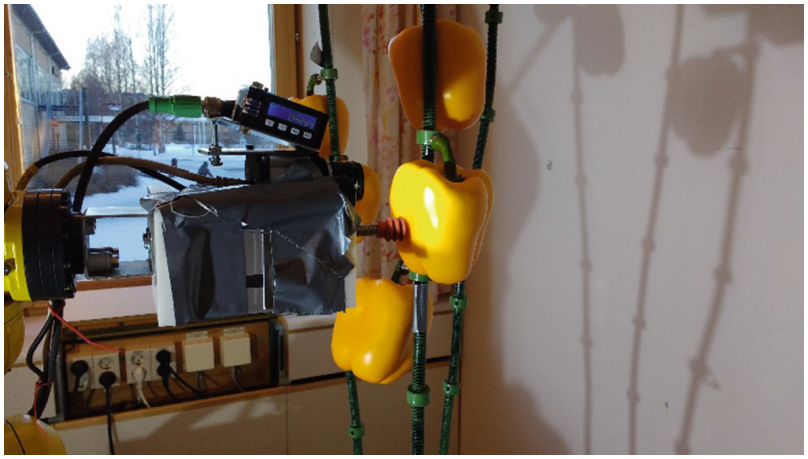


Fig. 5. The pose of each pepper was measured by manually moving the robotic arm in the desired approach direction to the position where the gripper touched the fruit.

A harvesting cycle is performed for each of the defined scenes and scenarios according to the following configurations. Each one of the scenes defined is performed in three possible configurations:

- Full a-priori knowledge scenario selecting the optimal approach direction from the set {−45°, 0°, 45°}
- Auto approach direction search scenario with two different search patterns:
  - Side first: $\theta_j = [-45°, 0°, 45°]$ (left-center-right)
  - Center first: $\theta_j = [0°, -45°, 45°]$ (center-left-right)

Each configuration is performed at 50% and 100% of maximum speed respectively to enable sensitivity analysis in relation to the robot speed. At the end of each harvesting attempt cycle times and the result of the attempt (success/failure) are registered.

2.3 Measures and Statistical Analysis

To evaluate the performance, the following three measures are defined:

- *Pepper harvest time Th* is the time it takes from a fruit is selected from the list of fruit poses until the fruit has been successfully harvested (all fruits were harvested in the experiments).
- *Average logarithmic harvest time LTh* as shown in Eq. 1.

$$LTh = \frac{1}{n}\sum_{i=1}^{n} \ln(Th_i) \tag{1}$$

Where n is the number of successfully harvested fruits.

- The *number of attempted approach directions* $N\theta_i$ for fruit i.

In addition to descriptive statistics of the aforementioned measures, the statistical significance of the differences in the value of the measures was measured. The pepper harvest time *Th* is analyzed in a form of a log transformed linear regression [16]:

$$\ln(Th_i) = \beta_0 + \beta_1 H_{c_i} + \beta_2 O_i + \beta_3 V_R + \beta_4 O_{F_i} + \beta_5 H_{c_i} * O_i + \epsilon_i \tag{2}$$

Where H_{c_i} is the harvesting scenario of pepper i, O_i is the number of occluding leaves, V_R is the robot speed, O_{F_i} is the front occlusion (1 if the front is occluded, 0 otherwise), and β_0, β_1, β_2, β_3, β_4, β_5 the corresponding weights of the regression to be estimated. Additionally, independence χ^2 test [17] is performed for analyzing the relation between the number of failed approach directions $N_{\theta F_i}$ and the harvesting scenario H_{c_i}.

3 Results

To determine the value of an optimal harvesting approach direction, a total of 180 fruit harvesting attempts were performed on 6 scenes with 5 artificial peppers each, in a set up according to Table 1, with different harvesting scenarios (full a-priori, center first search pattern, and side first search pattern) using two different robot velocities (50% and 100% of maximum). The total average harvest time $\overline{Th}$ for all combinations was 8.56 s (SD = 3.88). The distribution among the three harvesting scenarios is presented in Fig. 6. The results show roughly 40–45% increase in average harvest time when no a-priori information of the correct harvesting direction is available.

Homogeneous subsets Tukey-HSD test show a significant (p-value = 0.011) differences between *LTh* (Eq. 1) calculated from the full a-priori and the center first search pattern harvesting scenarios. The difference between *LTh* for full a-priori and side first search pattern harvesting scenarios was also significant (p-value = 0.006). The

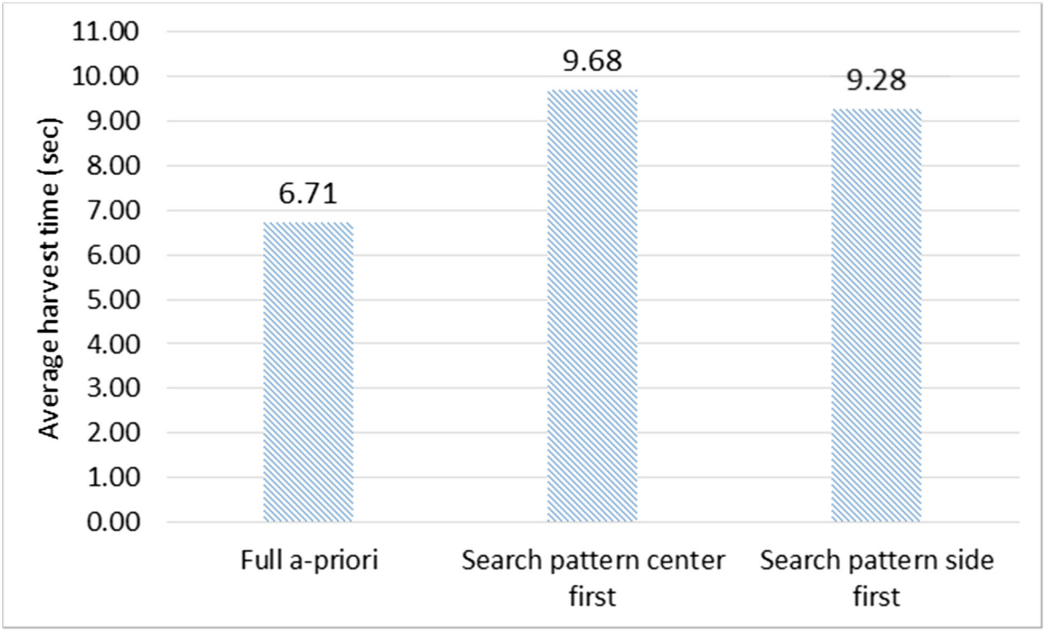


Fig. 6. Average harvesting time as function of the harvesting scenario

differences between *LTh* for the two search patterns were found to be statistically insignificant (p-value = 0.98).

Results of the logarithmic transformed $ln(Th)$ regression model (Eq. 2) revealed significance for front occlusion (p-value < 0.001) and harvesting scenario (p-value = 0.02). The number of occluding leaves was not found significant (p-value = 0.774) on its own but was borderline significant in an interaction with the harvesting scenario (p-value = 0.098). A profile plot describing the interaction is presented in Fig. 7. It shows that both search patterns have shorter harvesting times for less occluded scenes. It seems that in the full a-priori information scenario it takes slightly less time to harvest in more complicated scenes with higher occlusion then for simpler scenes. However, this difference was found statistically insignificant (p-value = 0.16). The difference between the two robot velocities (50% or 100% of maximum) was found to be insignificant (p-value = 0.155). This can be explained by the visual servoing technique that limits step sizes between images causing the robot not to obtain the maximum speed during this phase. This is needed to provide sufficient time to process image data during visual servoing.

From the total of 180 harvesting attempts performed, all 60 approaches (100%) performed with full a-priori information were successful on the first attempt with an average harvesting time of 6.71 s (SD = 3.05). Out of the 120 cycles performed using a search pattern, 76 (63%) were successful on the first attempt with average harvesting time of 6.62 s (SD = 2.78). 30 cycles (25%) were successful on the second attempt with average time of 11.16 s (SD = 5.4) and the remaining 14 cycles (12%) were successful only on the third attempt with average time of 21.34 s (SD = 6.9). The number of highly occluded peppers and partially occluded peppers were roughly the same (46% and 54% respectively). While the average harvesting time increased as a nearly linear function of the number of attempts, the standard deviation also increased for more complex cases requiring more attempts until harvesting. The analysis of the

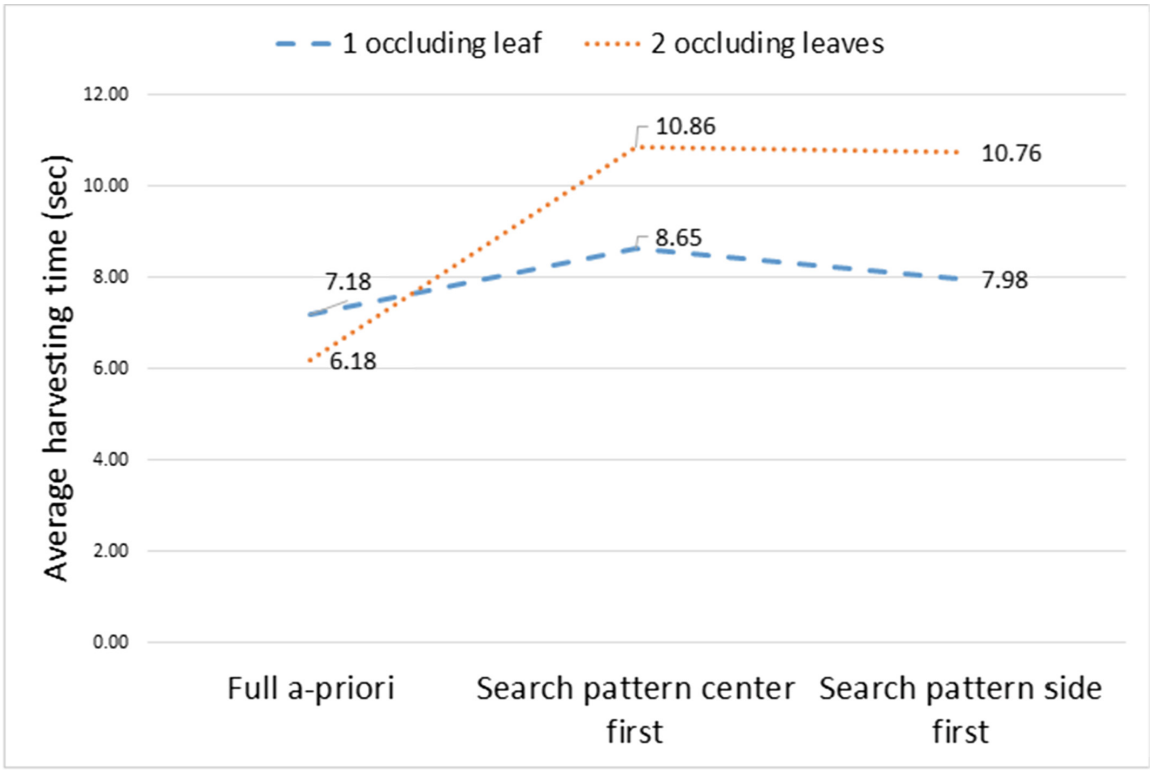


Fig. 7. Profiles plots for occlusion level and search method

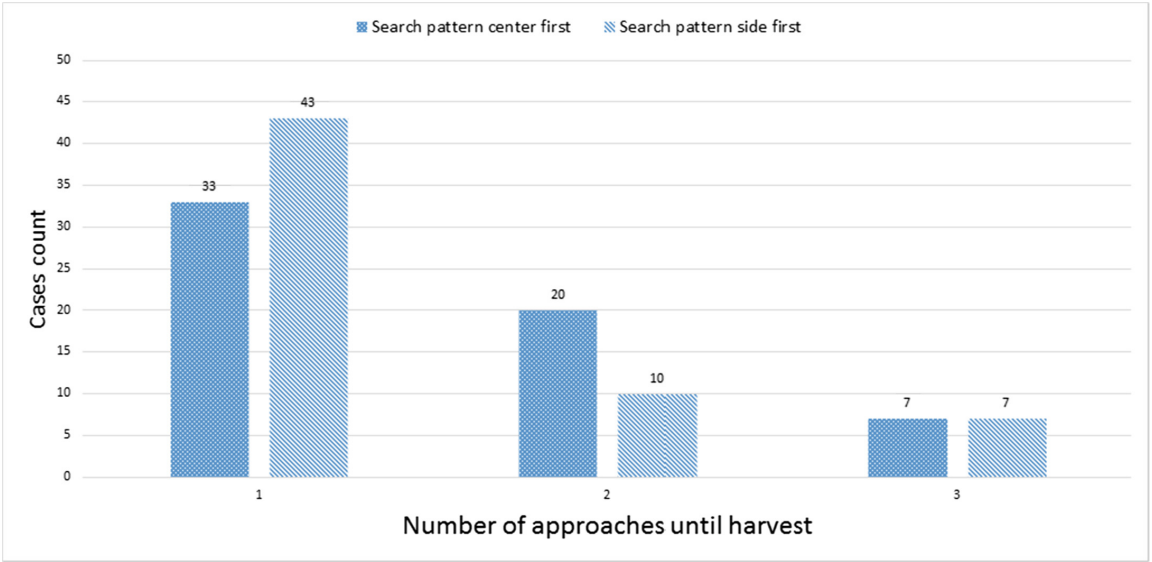


Fig. 8. Number of approaches until successful harvest as function of the search pattern method

number of approaches performed until successful harvest as function of search pattern method is presented in Fig. 8. It can be seen that about 30% more fruits were harvested at the first attempt using the side first search pattern than the center first pattern. An independence χ^2 test showed border line significant dependences between the search methods and the number of attempts (p-value 0.0978).

4 Conclusions

Results show significant increase in harvesting times for a search pattern compared to ideal initial information about the harvesting direction. The harvesting time grows near linearly with the number of approaches required until successful harvest. Furthermore, the variability of the harvesting time grows with the number of approaches required, causing lower ability to predict harvesting times. Therefore, it is clear that ideal information about the best harvesting approach direction is valuable for increasing the performance of a robot harvesting system.

The harvesting time does not significantly differ for the two different harvesting direction search patterns. This should be validated on a greater variation of search patterns and in greenhouse conditions where the occlusion is less likely to appear in a random manner as designed in the given experiment. To see how this depends on the kind of robot used, validating the results using a robot with different kinematic setup would also be beneficial. It has been shown that if there is an occlusion of the front of a fruit the harvesting times significantly increase compared to fruits that can be harvested from front, regardless of search method. The major reason for this is the limited workspace of the robot; the distance to the fruits is around 35–40 cm, with leaves often being even closer, and the gripper mounted on the end of the robot is 24 cm long. This makes it difficult to reach positions to the side of the peppers and the paths often become quite long due to the limited space and the joint limitations of the robot. Pruning techniques used for crops optimization might take this into consideration to facilitate robotic harvesting.

30% more fruits were harvested at the first attempt when using the side first search pattern than when using the center first pattern. Equal number of scene configurations had fruits blocked by leaves from left and center, therefore the number of approaches would have been expected to be equal for both search patterns. A probable explanation is that some fruits were detected during visual servoing even though they were (partly) blocked by leaves and therefore should not have been possible to harvest. This occurred in 26% of all attempts of harvest from the left and in 13% of all attempts from the front. However, this most likely did not affect the reported recall and precision since they are calculated in comparison to actual harvest approach success rates, i.e. that the robot actually reached the fruit.

The results of this research have shown significant factors affecting harvesting times and success rates in laboratorial conditions. Suggested validation of the results is to perform experiments in greenhouse conditions, which must be done during the growing season when ripe fruits are available.

Acknowledgments. This research was partially supported by the European Commission (SWEEPER GA no. 66313), by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Center, and by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, both at Ben-Gurion University of the Negev. The authors would like to acknowledge Peter Hohnloser at Computing Science department, Umeå University for his significant support and implementation of parts of the software system used in this research.

References

1. Comba, L., Gay, P., Piccarolo, P., Ricauda Aimonino, D.: Robotics and automation for crop management: trends and perspective. In: International Conference Ragusa SHWA 2010, pp. 471–478 (2010)
2. Bac, C.W.: Improving obstacle awareness for robotic harvesting of sweet-pepper (2015)
3. Bac, C.W., Henten, E.J., Hemming, J., Edan, Y.: Harvesting robots for high-value crops: state-of-the-art review and challenges ahead. J. F. Robot. **31**, 888–911 (2014)
4. Edan, Y., Flash, T., Peiper, U.M., Shmulevich, I., Sarig, Y.: Near-minimum-time task planning for fruit-picking robots. IEEE Trans. Robot. Autom. **7**, 48–56 (1991)
5. Harel, B., Kurtser, P., Van Herck, L., Parmet, Y., Edan, Y.: Sweet pepper maturity evaluation via multiple viewpoints color analyses. In: CIGR-AgEng Conference (2016)
6. Hemming, J., Ruizendaal, J., Hofstee, J.W., van Henten, E.J.: Fruit detectability analysis for different camera positions in sweet-pepper. Sensors **14**, 6032–6044 (2014)
7. Gongal, A., Amatya, S., Karkee, M., Zhang, Q., Lewis, K.: Sensors and systems for fruit detection and localization: a review. Comput. Electron. Agric. **116**, 8–19 (2015)
8. Vitzrabin, E., Edan, Y.: Adaptive thresholding with fusion using a RGBD sensor for red sweet-pepper detection. Biosyst. Eng. **146**, 45–56 (2016)
9. Rosenbaum, D.A., Cohen, R.G., Meulenbroek, R.G.J., Vaughan, J.: Plans for grasping objects. In: Motor Control and Learning, pp. 9–25. Kluwer Academic Publishers, Boston (2006)
10. Eizicovits, D., Berman, S.: Efficient sensory-grounded grasp pose quality mapping for gripper design and online grasp planning. Rob. Auton. Syst. **62**, 1208–1219 (2014). doi:10.1016/j.robot.2014.03.011
11. Eizicovits, D., van Tuijl, B., Berman, S., Edan, Y.: Integration of perception capabilities in gripper design using graspability maps. Biosyst. Eng. **146**, 98–113 (2016). doi:10.1016/j.biosystemseng.2015.12.016
12. Kapach, K., Barnea, E., Mairon, R., Edan, Y., Ben-Shahar, O.: Computer vision for fruit harvesting robots–state of the art and challenges ahead. Int. J. Comput. Vis. Robot. **3**, 4–34 (2012)
13. Barth, R., Hemming, J., van Henten, E.J.: Design of an eye-in-hand sensing and servo control framework for harvesting robotics in dense vegetation. Biosyst. Eng. **146**, 71–84 (2016). doi:10.1016/j.biosystemseng.2015.12.001
14. Hellström, T., Ringdahl, O.: A software framework for agricultural and forestry robots. Ind. Robot. Int. J. **40**, 20–26 (2013). doi:10.1108/01439911311294228
15. Ringdahl, O., Kurtser, P., Barth, R., Edan, Y.: Operational flow of an autonomous sweet pepper harvesting robot. In: The 5th Israeli Conference on Robotics 2016, Air Force Conference Center Hertzilya, Israel, 13–14 April 2016
16. Benoit, K.: Linear regression models with logarithmic transformations. London School of Economics, London (2011)
17. Greenwood, P.E., Nikulin, M.S.: A Guide to Chi-Squared Testing. John Wiley & Sons, New York (1996)

תקציר
המחקר מתמקד בשיפור ביצועים אופרטיביים של רובוט קטיף אוטונומי תוך התמקדות בחישה דינמית ותכנון משימות. השיפורים מובילים להגדלת אחוזי זיהוי והקטנת זמני קטיף – שני מדדים מרכזיים של רובוטי קטיף אוטונמיים, המונעים מהם כניסה לשוק. במסגרת המחקר הוגדרו ארבעה אתגרים מרכזיים – ניתוח נקודות מבט וחישה, חישה דינמית, תכנון משימות קטיף וגישה תכננון גישה לפרי. האלגוריתמים שפותחו נבחנו על מקרי בוחן של קטיף אוטונמי של פלפל. במסגרת התזה האתגרים נפתרו בין היתר על ידי איסוף מסדי נתונים נרחבים דרך פרוטוקלי איסוף נתונים חלקאיים בסביבת חממות. נאספו מסדי נתונים בקנה מידה גדול שיתוף עם חוקרים נוספים כחלק מהפרוייקט האיחוד האירופאי SWEEPER בו גם יושמו חלקית מסקנות מחקר זה.

ניתוח נקודות מבט וחישה. ניתוח סטטיסטי של נקודות מבט המאפשר תובנות ובחירת נקודות מבט מיטביות לביצוע חישה עבור זיהוי הפרי. נמצא כי מנקודת מבט בודדת ניתן לזהות 40-60% מהפרי. נמצאו תובנות לנקודות מבט מייטביות והחלטה כי שתי נקודות מבט הינן תנאי מינימלי להשגת זיהוי של עד 85% מהפרי בשטח.

חישה דינמית. מספר נקודות החישה בפועל צריך להיות מינימלי להקטנת זמן הקטיף. על כן פותח. אלגוריתם תומך החלטה לביצוע נקודת חישה נוספת אשר חוזה את מספר הפלפלים העתידים להיות מזוהים ומשווה את עלות איבודם אל מול עלות זמן ריצה ארוך יותר. האלגוריתם הראה שיפור הן בזמני הקטיף (בעד 10%) והן באחוזי הזיהוי (ב19%).

תכנון משימות הקטיף. במצב בו הפירות מזוהים אך ורק מתוך פעולות חישה ואינם ידועים מראש תכנון סדר החישה והקטיף נדרש לשיפור זמני הקטיף. בעבודה זו פותחו אלגוריתמים ליעול סדר הפעולות אשר הובילו להקטנת זמני המחזור בממוצע ב12%.

תכנון גישה לפרי. יש חשיבות לגישה לפרי מנקודת מבט אשר אינה חסומה על ידי עלים. במסגרת המחקר נבחנו מספר שיטות לבחירת נקודת מבט זו ונבחנה יכולת האדם לזהות נקודות גישה מייטביות.

הצהרת תלמיד המחקר עם הגשת עבודת הדוקטור לשיפוט

אני החתום מטה מצהיר/ה בזאת: (אנא סמן):

✓ חיברתי את חיבורי בעצמי, להוציא עזרת ההדרכה שקיבלתי מאת מנחה/ים.

✓ החומר המדעי הנכלל בעבודה זו הינו פרי מחקרי מתקופת היותי תלמיד/ת מחקר.

תאריך 10.10.2018 שם התלמיד/ה פולינה קורצר חתימה P.K.

העבודה נעשתה בהדרכת פרופ׳ יעל אידן
במחלקה להנדסת תעשיה וניהול
אוניברסיטת בן-גוריון בנגב

תכנון משימות וחישה דינמית עבור רובוט קטיף אוטונומי

מחקר לשם מילוי חלקי של הדרישות לקבלת תואר "דוקטור לפילוסופיה"

מאת

פולינה קורצר

הוגש לסינאט אוניברסיטת בן גוריון בנגב

אישור המנחה _____

אישור דיקן בית הספר ללימודי מחקר מתקדמים ע"ש קרייטמן _____

א' חשוון תשע"ט **10.10.2018**

באר שבע

תכנון משימות וחישה דינמית עבור רובוט קטיף אוטונומי

מחקר לשם מילוי חלקי של הדרישות לקבלת תואר "דוקטור לפילוסופיה"

מאת

פולינה קורצר

הוגש לסינאט אוניברסיטת בן גוריון בנגב

א' חשוון תשע"ט **10.10.2018**

באר שבע