Project title:	Soft fruit detection and shape estimation using 3D information and machine learning
Project number:	SF/TF 170
Project leader:	Grzegorz Cielniak, University of Lincoln Charles Whitfield, NIAB EMR
Report:	Annual report, 2019
Previous report:	NA
Key staff:	Justin Le Louedec
Location of project:	University of Lincoln
Industry Representative:	Berry Gardens Growers
Date project commenced:	September 2018

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

AUTHENTICATION

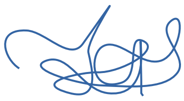
We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name] Justin Le Louedec

[Position] PhD student

[Organisation] University of Lincoln

Signature Date .14/01/2020



Report authorised by:

[Name] Grzegorz Cielniak

[Position] Supervisor

[Organisation] University of Lincoln

Signature Date 14/01/2020

TABLE OF CONTENTS

GROWER SUMMARY **5**
Headline 5
Background 5
Summary 5
Financial Benefits 5
Action Points 5
SCIENCE SECTION **6**
Introduction 6
Materials and Methods 7
Results 8
Discussion and Conclusion 12
Knowledge and Technology Transfer 13

GROWER SUMMARY

Headline

We aim to study the utility of 3D information for robotic perception in commercial strawberry production. We have found, through investigating sensing technologies and state-of-the-art algorithms for the detection of fruit that using images rather than 3D information is more suitable. We will now focus on shape description and understanding, as well as 3D reconstruction of fruit.

Background

This work is part of a larger programme of work funded by BBSRC and industry through the Collaborative Training Partnership for Fruit Crop Research (CTP-FCR) to develop and deploy robotic platforms for commercial strawberry production. The platform is hoped to assist pickers, growers and agronomists to deliver tasks. in this current study we will develop the vision system component of the robotic platform and test how it can be utilised for different tasks such as picking and phenotyping.

Summary

During the first year of this project we have studied the sensing technologies available and state-of-the-art algorithms for the task of detecting strawberry fruit in 3D space. The results have developed methods of understanding shape and phenotyping of the berries.

Financial Benefits

This project is part of a much larger programme to develop robotics for the horticultural industry. The exact financial outcomes of such investment in robotics and computer science is unclear at this early stage. However, it is expected that a fully working robot picker would alleviate labour cost for picking, transporting and analysing fruits in the grower facility, with an initial investment in the robot.

Action Points

There are no action points at this early stage of the project

SCIENCE SECTION

Introduction

Agriculture is rapidly becoming an important industry for the application of robotics technology and artificial intelligence. With a growing population, climate change and a shortage of cultivatable land, new solutions need to be found to provide food in an efficient and clever way. While a lot of work is done in research and industry, research for new and innovative techniques and solutions has never been as important. Robots and artificial intelligence perception algorithms are key to improving production yield and product quality in the agricultural industry.

While algorithms have been created to take advantage of 2D images to detect fruits and vegetables, there is a lack of interest given to spatial information found in 3D data. This can be explained by the difficulty to work with this type of data and the lack of affordable and precise sensors to capture 3D information for outdoor applications. RGB-D (colour and depth) cameras are a huge leap forward, providing more information (combining the colour to the geometrical information) about the environment. However modern algorithms taking advantage of 3D information tend to be task specific and have difficulties to adapt to new environments or goals.

We have seen a lot of advancement in computer vision concerning Deep Learning for object detection or instance segmentation and this technology has been applied to agricultural challenges such as soft fruit or vegetables picking. Deep Learning is part of the broader field of machine learning and about creating algorithm learning representations and tasks from a large amount of data. These algorithms are inspired by biological information processing systems from a structural point of view. Using a succession of layers or so-called "neurons", information is extracted from the data as feature maps and processed depending on the goal, to create a generalised representation of our data.

RGB-D cameras give us the 3D information as a depth map that can be projected into a 3D space using intrinsic and extrinsic parameters. This projection is called a point cloud, which is a simple and unified structure with some constraints (unordered set of points in Euclidean space, but with interaction among those points such as geometry or surface normal). This representation brings more information than images, such as localisation, shape or size of objects for example.

Creating Deep Learning algorithms for this type of representation, used in an agricultural context with visual data directly for fruit picking would; help in yield prediction; provide more data to agronomists and researchers; and improve production and picking of fruits such as strawberries. Also, such algorithms will not be dependent on the type of fruit or vegetables and could through its training process, be used for various tasks and products.

It is of the highest importance of developing a fast and reliable Deep Learning algorithm to detect fruits using 3D information and infer shape, size or other aspects of the fruit. However, our preliminary work, highlighted the difficulty to achieve the same performances in 3D as the one achieved using more mature techniques for 2D? images such as…. This is due to the difficulty to capture 3D information in an outdoor agricultural context and why we decided to rely on mature and reliable convolutional machine learning techniques for the task of detection, and to focus instead on shape description and analysis using 3D information. We aim at providing reliable shape descriptor, both handcrafted and learned, that can be used for shape analysis, reconstruction and phenotyping of strawberries or other soft fruits. We aim at proposing algorithms extracting shape descriptors from point clouds and reconstruct using spectral domain such as spherical harmonics or Laplacian. Such algorithms allow us to represent objects in a latent space, that can be then processed to extract information such as size, volume or circumference, but also can help reconstructing the object after degradation or seen from a single view. Description of shape (phenotype), associated with other information such as genotype or environment information, can help discover interesting correlations or predictions for yield, production, maturity or evolution. Using conclusions from state-of-the-art work and our preliminary work on point clouds and graphs we now work at improving current state-of-the-art for point cloud and graphs to provide understandable results that can be easily read by humans and related to other fruit attributes.

Materials and Methods

During this first year of PhD we worked on the study of 3D sensing technology and state-of-the-art algorithms to process such information for fruit detection. 3D sensing is based on capturing the depth or distance from the camera to each point in the scene. Capturing devices for outdoor use can be divided into three main categories: stereo cameras, Time-of-Flight (ToF) devices and Lidar range finders. Stereo sensors are based on capturing two images from two image sensors apart from each other and matching their features to create a depth map based on epipolar lines between the two sensors. In the case of wrongly matched points or a lack of similarity, surfaces reconstructed in this way, are often distorted or flat with blended edges and objects. This is especially evident with very small objects. Alternative sensing solutions use light wavelengths outside of the visible spectrum (e.g. infra-red) which are less prone to changing lighting conditions and more robust matching points. Time-of-Flight devices are based on light beams which are being projected into the scene and reflected back to the sensor. The depth is estimated from the time taken for the light to come back. This technology results in more precise depth measurements, but more prone to noise caused by reflective objects. The Microsoft Kinect One (i.e. v2) and the Pico Zense are a good example of recent innovations in this technology. Lidar is a particular example where the beam of light is replaced by a laser pulsed at the scene. We do not consider Lidar technology in our work, however, since its intrinsic properties and resolution are not suited for the detection of small objects in occluded scenarios such as strawberries.

This year, we compared stereo and Time-of-Flight sensing technologies based on their performance in sensing of strawberries in their natural growing conditions. The two selected cameras were the Intel Realsense D435 (IR stereo) and the Pico Zense (ToF).

We also worked on modern 3D algorithms and on their performances compared to more mature 2D algorithms. We were interested in the feasibility of modern 3D sensing and machine learning for a problem of detecting strawberries in their natural environment.

We also worked on collecting data to support the main goal of our application, which is applying a modern 3D vision system for the detection and localisation of strawberry fruit, we collected a dataset from the real environment. To that end, we have deployed our data acquisition system at a facility located at Riseholme campus of the University of Lincoln that simulates commercial strawberry production. The plot features two polytunnels of 6 tabletop rows, 24 meter long with an commercially relevant variety of strawberry (everbearer Driscoll's Amesti).

The data capture setup featuring the Realsense and Pico Zense sensors was mounted on an agricultural robot Thorvald (with the assistance of the other computer science student of the project working on strawberry prediction and yield forecasting, Raymond Kirk). The robot autonomously navigated the polytunnel rows, stopping every 20 cm to collect a snapshot from both views. The capturing session took place in October 2019 and resulted in colour images and point clouds representing different growth stage of plants and fruit. The datasets were then manually annotated to indicate location of strawberry fruits resulting in 139 labelled point clouds with around 1900 instances of ripe strawberries for ToF data and 64 point clouds for around 1000 instances for stereo data.

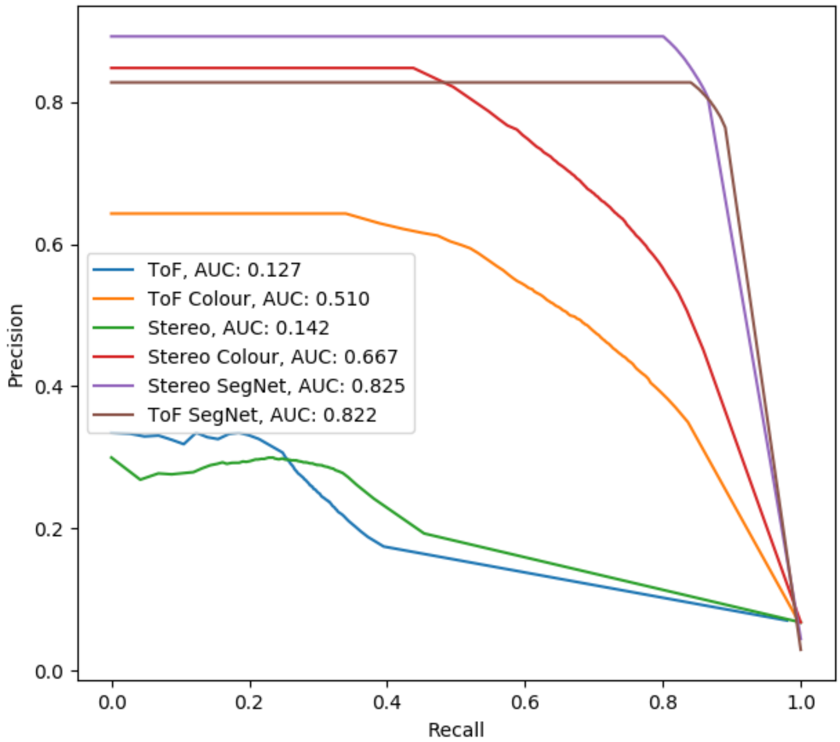
Results

To compare both sensors directly, we use the captured data for training the selected machine learning networks for each sensor separately. In addition, we use the following configurations of the networks and type of input data: SegNet on colour images, PointNet++ on 3D point clouds only (PNet) and PointNet++ with additional colour information. The results are presented in Table 1 and the precision-recall curves are presented in Figure 1.

Table 1: Comparison of different networks trained on data for stereo and ToF sensors. IoU refers to the overlapping between annotation and prediction

Model	**Camera**	**Dimension**	**Accuracy**	**Kappa Cohen**	**Mean IoU**
SegNet	Stereo	2D	98.5%	0.71	0.83
SegNet	ToF	2D	99%	0.79	0.67
PNet	Stereo	3D	91.1%	0.43	0.14
PNet	ToF	3D	90%	0.38	0.15
Pnet and colour	Stereo	3D	95.4%	0.66	0.48
Pnet and colour	ToF	3D	92.4%	0.54	0.39

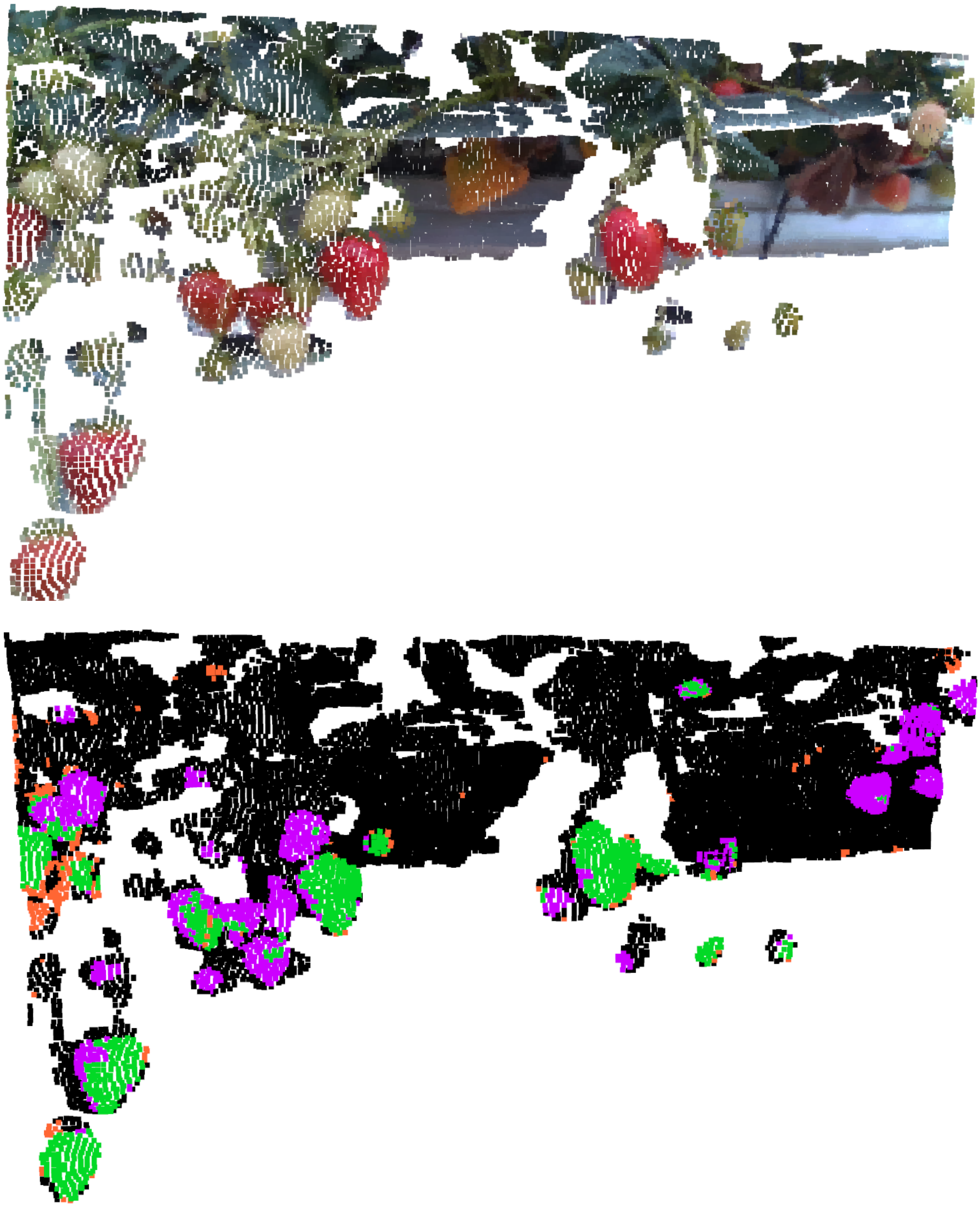
Figure 1: Precision Recall curve for the trained networks of Table 1



There is a clear difference between networks trained with and without colour information and between networks trained on ToF and stereo datasets with performances significantly improved for both datasets. This can be explained by a greater difference using colour space between strawberries and background than using only shape information. The difference in results for different sensors seems to be amplified when colour information is used and would come from the unreliable readings from large surfaces and shapes with ToF cameras leading to many false positives (FP). The precision-recall curves confirm these findings. The high amount of FP is negatively affecting the networks trained with 3D information only but improved significantly for networks trained in colour. With a very low area under the precision-recall curve, $PNet$ is the worst performing classifier on our datasets. A 2D network SegNet performs significantly better than any of the 3D variants for both datasets. To illustrate these findings, we provide example outputs from PointNet with colour illustrating the quality of segmentation for both sensors. The stereo dataset (Figure 2) is characterised by more omissions of strawberries with a high number of FN, but less false detections. The main red and distinct strawberries are however well segmented.

The superior performance of detectors based on 2D CNNs for our application can be associated with the structured nature of 2D images, maturity of the developed networks and also low quality of the depth data. Also, through post-processing, sufficient spatial information can be retrieved using the depth map and the 2D segmentation mask, making these algorithms preferable for our application at the time being. The presented 3D approaches, however, offer an advantage in direct localisation of the objects, although their localisation accuracy is a subject of future work. The real-time suitability of the 3D methods is also promising, achieving 5 FPs (each frame a point cloud of ~64k points) compared to 13 FPs (each frame an image of 1280 * 720 px) for SegNet.

Figure 2: Results obtained using PointNet with colour on the stereo data. The colours indicate: TP in green, FP in orange, FN in purple and TN in black.



Discussion and Conclusion

Capturing 3D data and processing it for different tasks such as detection, segmentation or classification is a challenging task especially in the agricultural context presented in this study. Our study evaluated two 3D sensing technologies for that purpose and compared 3D and 2D variants of state-of-the-art neural networks trained on the data collected from a commercial strawberry setting. These results show encouraging performance but also allow us to highlight the limitations of current technologies and algorithms. Time-of-Flight technology, despite its superior quality of point clouds and shape information, struggles with reflective surfaces resulting in a large number of false detections, while stereo technology, lacking detail in acquired depth, fails to detect numerous fruits. Traditional 2D image-based convolutional neural networks still outperform the 3D networks for the task of fruit segmentation and therefore are more suited for this task. This work can be treated as a baseline for future work on 3D information for outdoor applications such as robotic fruit picking. Researchers should be encouraged to pursue more experimentation in such challenging conditions to counteract limitations found in this study and bridge the gap with state-of-the-art techniques in perception for 2D information.

Based on the literature we have identified the need for using 3D shape analysis and reconstruction for applications in agriculture such as yield prediction or picking. The current state-of-the-art technologies are being tested on datasets very different from the conditions we find in agriculture and are using existing datasets. We are dealing with challenging conditions and also needed to create our datasets. The important aspect in 3D is the shape of the objects in the scene, which make them unique, particularly for strawberries and fruits. Investigating a new way to capture shape characteristics of the objects and analysing it, is of the highest importance.

We have captured a dataset of strawberries in the field suiting our needs and annotated it, enabling us to determine a baseline using state-of-the-art algorithms on this dataset. We are now able to identify the shortcomings and challenges with this application and these algorithms. The next step is to move on to shape descriptors and analysis. For that, we will start with standard approaches on high-resolution data captured in a controlled environment before moving on to learned descriptors and comparing them. We will then move toward studying if these descriptors are transferable on more difficult and less complete datasets captured in the field. Finally, we wish to create the link between these vision techniques and intrinsic characteristics of fruits such as size and volume for phenotyping and genotyping.

Knowledge and Technology Transfer

Events attended:

- BMVA symposium “Deep Learning in 3 Dimensions” in London, February 2019
- Project workshop at NMBU (Norwegian University of Life sciences), December 2018 and May 2019
- CTP meetings :
    - Autumn 2018
    - Spring 2019
    - Autumn 2019
- 15th International Conference on Vision Theory and Applications, February 2020
- AHDB crops Studentship Conference, January 2020

Technical skills improvement:

- Improvement in scientific writing through multiple reports.
- Improvement in presentation skills, by participating in various events, symposiums and conferences.
- Poster creation and presentation during events.